AR-010-542

ATM Signalling in Perspective

T. A. Au

DSTO-TR-0671

DEPARTMENT OF DEFENCE
DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

AQF98-12-2382

ATM Signalling in Perspective

T. A. Au

Communications Division
Electronics and Surveillance Research Laboratory

DSTO-TR-0671

ABSTRACT

Signalling has great impact on the efficient use of network resources, and the services that an Asynchronous Transfer Mode (ATM) network can offer. Importantly, signalling allows the transfer of service-related information in real time between the user and the network, and among network entities to establish, maintain and release end-to-end virtual connections. For each particular signalling function, the corresponding procedures are defined such that the sequence and message format are specific to the network interface across which the exchange of signalling information takes place. This report reviews the signalling architectures for the user-network interface and the network-network interface in both private and public ATM network environments. It is hoped that readers will gain a proper perspective of the signalling protocols involved in ATM networks, and thereby an appreciation of their significant contributions.

RELEASE LIMITATION

Approved for public release

DEPARTMENT OF DEFENCE

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

Published by

DSTO Electronics and Surveillance Research Laboratory
PO Box 1500
Salisbury South Australia 5108

Telephone: (08) 8259 5555
Fax: (08) 8259 6567

AR-010-542
May 1998

APPROVED FOR PUBLIC RELEASE

ATM Signalling in Perspective

Executive Summary

In the evolution of telecommunications networks, Asynchronous Transfer Mode (ATM) technology is rapidly becoming the universal networking standard as an approach to integrating all different kinds of communication traffic. Typically, ATM networks operate in connection oriented mode in which a pre-established virtual circuit is required for every connection. Signalling, as an essential component, has great impact on the efficient use of network resources, and the services that a network can offer. It is important that the details of a connection can be specified and conveyed to the network, otherwise the user will not be able to receive the required services in a dynamic manner. Essentially, signalling allows the transfer of service-related information in real time between the user and the network, and among network entities to establish, maintain and release end-to-end virtual connections.

The standardisation activity on signalling protocols is being undertaken by the ITU-T and the ATM Forum to facilitate a widespread deployment of ATM services. Two ATM interfaces are defined for signalling purposes. The user-network interface (UNI) is the interface between an endpoint equipment and the network, whereas the network-network interface (NNI) is the interface between either two private networks or two public networks. The separation of the private and public domains is a consequence of the significant differences in the administrative responsibility. For each particular signalling function, the corresponding procedures are defined such that the sequence and message format are specific to the network interface across which the exchange of signalling information takes place.

This report reviews the signalling architectures for the user-network interface (UNI) and the network-network interface (NNI) in both private and public ATM network environments. It is hoped that readers will gain a proper perspective of the signalling protocols involved in ATM networks, and thereby an appreciation of their significant contributions. In recent years, the field of telecommunications has witnessed the rapid evolution of signalling to facilitate emerging applications. As ATM technology becomes more mature, the signalling architecture is progressively empowered to support additional capabilities, including multiconnection, multimedia, and multipoint services. It is the signalling mechanism that will make ATM more flexible, efficient and adaptable to many environments.

Authors

T. A. Au
Communications Division

Andrew Au is a Research Scientist in Network Integration Group with interests in traffic control and performance guarantees for high speed networks. He has been investigating ATM signalling and communications issues in distributed computing environments.

Contents

1. Introduction

The emergence of Asynchronous Transfer Mode (ATM) networks as a technology of very high-speed communications will play a central role in the evolution of telecommunications networks. The deployment of ATM services requires the overlay of a highly complex, software intensive, protocol infrastructure. This is the foundation joining individual ATM switches into a network, and on that basis, the potential of ATM backbones in interconnecting isolated local area networks has been witnessed in recent years.

Signalling is an important component of communications networks which has great impact on the efficient use of network resources, and the services that a network can offer. Signalling allows the transfer of service-related information in real time between the user and the network, and among network entities to establish end-to-end communications. An ATM network operates in connection oriented mode in which a pre-established virtual path is required for every connection. It is important that the details of a connection can be specified and conveyed to the network, otherwise the user will not be able to receive the required services in a dynamic manner. Therefore, ATM signalling is essential to dynamically establish, maintain, and release virtual connections. To date, most ATM networks are implemented by setting up permanent virtual paths. This temporary solution does not require signalling since the virtual circuits between communicating endpoints are set up via management control.

In recent years, the field of telecommunications has witnessed the rapid evolution of signalling to facilitate new applications in ATM networks. These signalling protocols allow the exploitation of the full potential of ATM and permit the emerging services that ATM networks can provide to users. Typically, the standardisation activity on signalling protocols is being undertaken by the ITU-T[1] for public networks and the ATM Forum for private networks. These international bodies have been working on standards and recommendations to facilitate a widespread deployment of interoperable ATM networks.

This report is intended to provide an introduction to ATM signalling, in both private and public network environments. Signalling architectures for the user-network interface and the network-network interface are reviewed. It is hoped that readers will thereby gain proper perspective of the signalling protocols involved in ATM networks. Nevertheless, interested readers are encouraged to refer back to the original specifications and recommendations by the ATM Forum and the ITU-T for detailed descriptions.

Following this introduction, Section 2 presents the basic transfer principles of an ATM network. Virtual path and virtual channel concepts are described in detail. Section 3

[1] International Telecommunications Union - Telecommunication Standardisation Sector (ITU-T), formerly CCITT.

defines various types of standard interfaces in an ATM network. Section 4 provides a brief overview of control signal functionality in a public telecommunications network. Section 5 discusses some key elements of the signalling used to dynamically set up and clear ATM connections. The signalling ATM adaptation layer to support signalling connections is described in Section 6. The user-network interface is explained in Section 7 including signalling procedures for basic point-to-point call and point-to-multipoint call. Section 8 discusses the network node signalling in private ATM networks. As a very important part of the evolving ATM standards, the private network node interface is described in detail. In the public network environment, Section 9 describes the signalling system number 7, the B-ISDN user part and the broadband inter-carrier interface. Section 10 provides a summary of signalling procedures in setting up an ATM connection. Future development of signalling protocols is discussed in Section 11.

2. ATM Data Transfer Principles

The transport function of an ATM network is structured hierarchically in which a 53-octet cell is the unit of switching and transmission. The transmission medium, or the physical link, provides transport service to the virtual path (VP), identified by a virtual path identifier (VPI). The VP in turn provides transport service to the virtual channel (VC), identified by the combination of a VPI and a virtual channel identifier (VCI). The relationship between a physical link, a virtual path and a virtual channel is illustrated in Figure 1 [1].

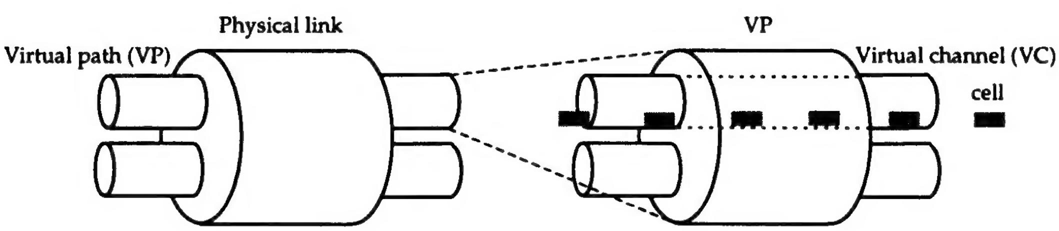


Figure 1: Physical Link, Virtual Path and Virtual Channel Relationship

To uniquely identify an ATM connection to which each cell belongs, the routing address is indicated by a VPI/VCI value in the 5-octet cell header. VPIs are uniquely defined on each transmission link and VCIs are uniquely defined on each virtual path. An ATM switch must keep a table of VPI/VCI values relating to each attached physical link. Incoming cells are transported based on the information stored in this routing table using the VPI/VCI value. An ATM switch retransmits the cells on an outgoing link with a new VPI/VCI value which determines an appropriate connection for the next switching element.

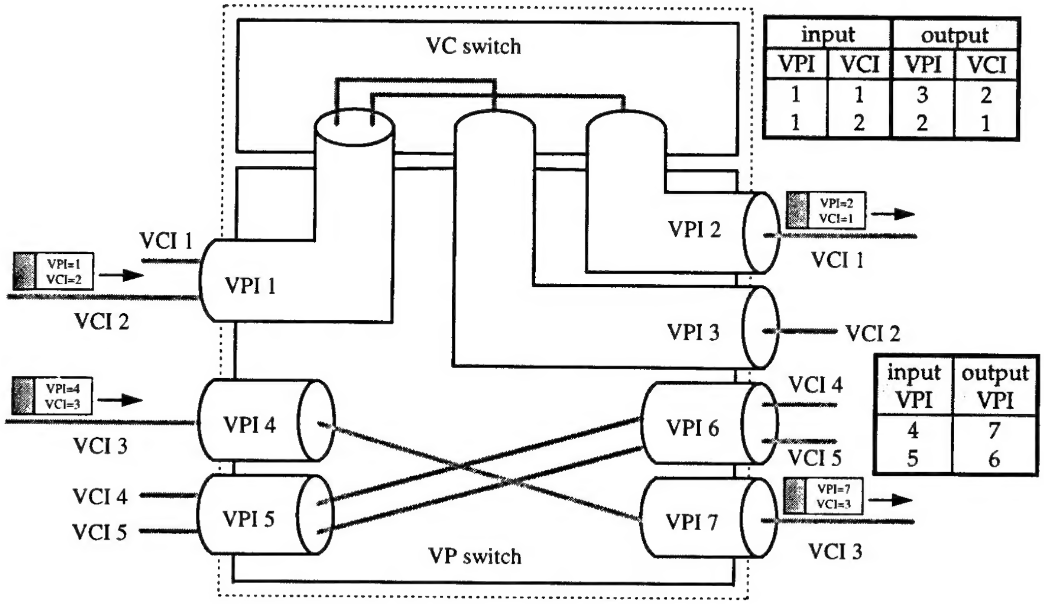


Figure 2: virtual channel and virtual path switching

As explained in Figure 2, two types of switches are defined in ATM networks: VP switch and VC switch. In VP switching, all of the VCs of a VP are switched transparently from the input link to an output link on the basis of the common VPI. A specific VPI value is assigned each time a VP is switched in the network. VCI values of the VP are unchanged and have no significance in VP switching. VC switching, on the other hand, routes the incoming cells by using both the VPI and the VCI fields. VPs are terminated at a VC switch and new values of the VPI and the VCI are assigned.

The VP and VC concepts create two kinds of connections in ATM networks: virtual channel connections (VCC) and virtual path connections (VPC). A VPC can be considered as an aggregate of VCCs. By definition, a VPC is a concatenation of VP links. A VP link provides unidirectional transport of ATM cells between the point where the VPI value is assigned and where it is translated or the VP is terminated. Likewise, a VC link is defined from the point where a VCI value is assigned to where it is translated or the VC is terminated. A concatenation of unidirectional VC links forms a VCC. These are illustrated in Figure 3.

ATM networks are fundamentally connection oriented in which the cell sequence on a virtual connection is maintained. An end-to-end virtual connection has to be established prior to any data transfer. External mechanisms are involved in the setup of routing tables in switches along the connection route. Connections are established either semi-permanently, or for the duration of a call, in case of switched services. These connections are referred to as permanent virtual circuits (PVCs), and switched virtual circuits (SVCs) respectively.

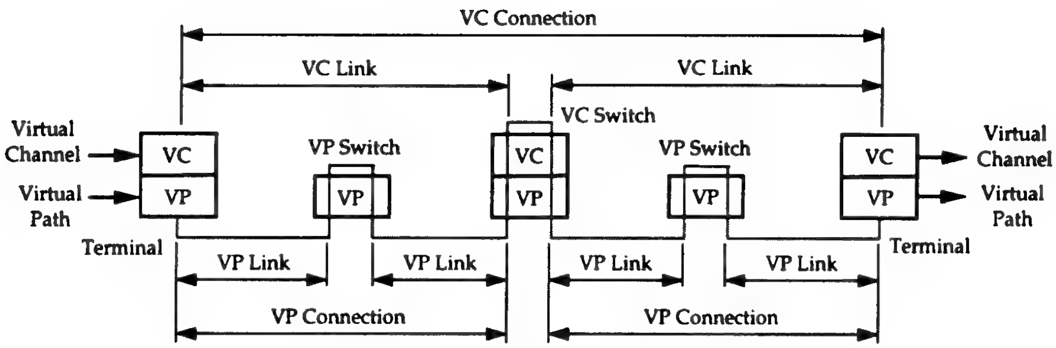


Figure 3: VP and VC concepts

- Permanent Virtual Circuit (PVC) - A PVC is a pre-established connection with static route defined in advance, usually by manual configuration. Through some network management functions (OAM procedures[2]), a set of switches between predefined endpoints are programmed with the appropriate VPI/VCI values. Analogous to a leased line, a PVC is always present and no call setup is required. Restoration of permanent ATM connections is automatic after network failure. In ATM standards, two kinds of PVCs are available. The first, a permanent virtual channel connection (PVCC) is a pre-established VCC where switching is performed on the VPI/VCI fields of each cell. The other, permanent virtual path connection (PVPC) is a pre-established VPC where switching is performed on the VPI field only of each cell.

- Switched Virtual Circuit (SVC) - An SVC is a connection that is set up dynamically on demand using signalling. The user defines the endpoints when the call is initiated. Unlike PVCs, establishing SVCs does not require manual interaction or network management functions. An SVC is lost through a network failure, and the user has to forward another request for further connection. SVCs are comparable to telephone calls in which connections are set up dynamically as needed and torn down immediately afterwards. A switched virtual channel connection (SVCC) is a switched ATM connection established via signalling where switching is performed on the VPI/VCI fields of each cell. Likewise, a switched virtual path connection (SVPC) is a switched ATM connection where switching is performed only on the VPI field.

3. ATM Network Interfaces

An ATM network consists of a set of ATM switches interconnected by point-to-point ATM links. Various types of interfaces have been defined by the ITU-T and the ATM

[2] OAM stands for Operations Administration and Maintenance. It represents a group of network management functions that provide network fault indication, performance information, and data and diagnosis functions.

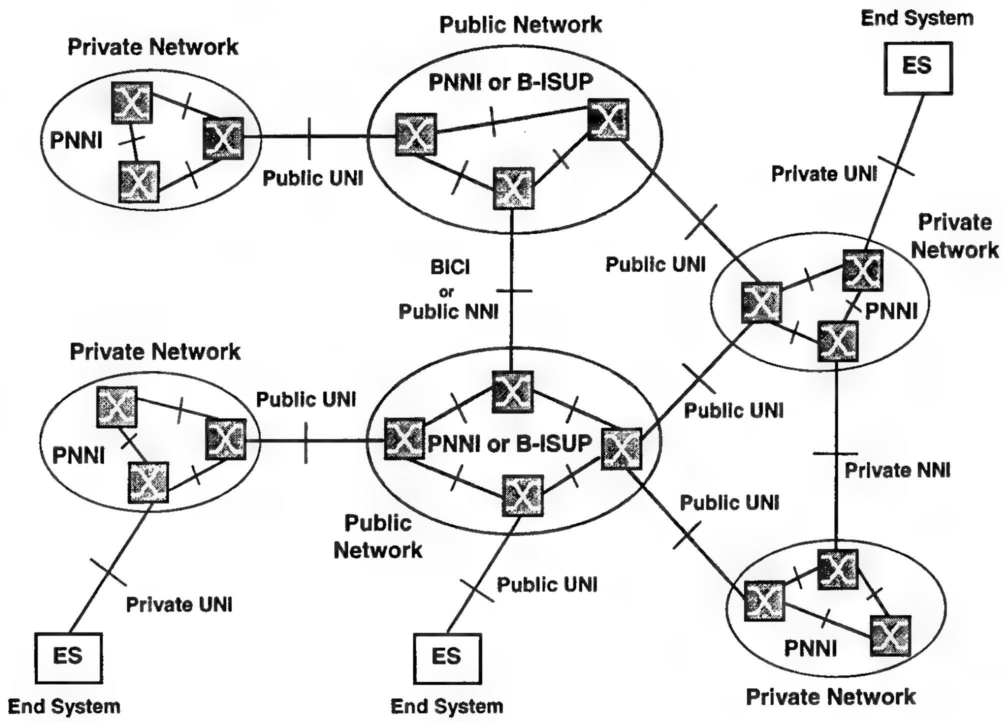


Figure 4: ATM Network Interfaces

Forum which address the connectivity and interoperability issues between the individual components of ATM networks. Figure 4 illustrates different types of interfaces in possible interworking scenarios. Interface protocols are specified in ATM standards to support end-to-end connectivity across globally connected ATM networks, including private and public networks.

Since the issues are completely different in public and private environments, public and private ATM networks are considered independently. Public ATM networks, also called service provider networks, are operated by public carriers. Private ATM networks, on the other hand, are exclusive to particular organisations who use, rather than provide communication services. Private ATM networks are often confined to local area such as a building or a campus. Extension of a private ATM network over a wide area has to make use of carrier links between ATM nodes. Public network equipment usually requires much higher throughput and more sophisticated network management procedures compared to private network equipment. ATM switches are therefore classified as either private or public ATM switches. An end system (ES), or an ATM endpoint, is a system where an ATM connection is initiated or terminated. It is a piece of end-user equipment that interfaces to an ATM network.

Information exchange is required between network users and the switching nodes, and between switching nodes. The user-network interface (UNI) is the interface between an ATM end system and the network where the end system or user can access the

network services. Depending on whether the network is private or public, the interface is respectively referred to as a private UNI or a public UNI. Typically, the interface between a private network and a public network is also called a public UNI because no routing information is exchanged. The requirements associated with the public and private UNIs make some functional differences. Connections between users and public ATM switches generally span longer distances. In contrast, private ATM switches are often located in close proximity to the user equipment and hence limited distance technologies can be used in a private UNI.

The network node interface[3] (NNI) is the interface between either two private networks or two public networks, respectively referred to as a private NNI (PNNI) and a public NNI. The NNI can also be used as the interface between ATM switches within a public or private network. In the context of public networks, the ITU-T standardisation on NNI includes the specifications of physical and ATM layers; user, and management planes; and NNI signalling. Public network signalling must reliably handle all network management functions and respond satisfactorily to call attempts at any time.

The PNNI is designed to support a less stringent environment than the public NNI. Moreover, the PNNI is specified by the ATM Forum in an attempt to define the internal operations of an ATM network. The PNNI includes two categories of protocols: the PNNI routing protocol and the PNNI signalling protocol. The PNNI routing protocol enables ATM switches from different vendors to be integrated in the same network, whereas the PNNI signalling protocol enables the establishment of point-to-point and point-to-multipoint connections across the ATM network. The PNNI automatically and dynamically distributes routing information, enabling any switch to determine a path to any other switch.

The interim inter-switch signalling protocol (IISP), formerly known as PNNI phase 0, is a limited hop-by-hop routing scheme. Manual configuration of the static topology and resource tables is required to establish SVCs between ATM switches. In fact, the IISP is just a modified UNI protocol for inter-switch signalling, with nodes arbitrarily taking the role of the network and user side across particular switch-to-switch links.

The B-ISDN inter-carrier interface (B-ICI) is an ATM Forum defined specification for the interface between public ATM networks to support user services across multiple public carriers. Similar to the PNNI, the B-ICI however excludes the routing protocol as carriers tend not to share detailed routing information with their competitors.

The broadband ISDN user part (B-ISUP) is the set of recommendations at the NNI specified by the ITU-T. It is based on the existing signalling system number 7 (SS7) which defines the signalling messages to control connections and services for B-ISDN applications. The B-ISUP contains a family of signalling procedures used to set up, manage and tear down connections as well as to exchange the associated information.

[3] Also known as the network-network interface.

Nevertheless, public carriers may implement proprietary protocols or use an existing standard in switch-to-switch signalling communications within a public network.

4. Overview of Signalling

In a traditional telecommunications network, signalling is regarded as the collection of procedures for network management and for the exchange of control information among network entities to dynamically establish, maintain, and terminate connections. A relatively complex control signalling scheme is necessary for a large network, and its functions can be roughly grouped into four categories [2]:

1. supervisory - supervisory control signals are used to determine the availability of the called user and the requested resources. These control functions have a binary character (true/false, on/off) that indicates the status of control signals such as request for service, answer, alerting, and return to idle.

2. address - address signals are generated by a caller which carry the address numbers of a destination in a network. The address identifies the called user and supports the routing function through the network.

3. call information - call information signals provide information to the caller about the status of a call. In a public telephone network, audible tones are often used to convey the progress in completing a call, such as a dial tone and a busy tone.

4. network management - network management signals are messages used to assist network administrators in monitoring and control of network elements for configuration, fault, performance, security, and accounting purposes.

There are two methods of carrying control information or signals over a telecommunications network, namely, in-channel signalling, and common channel signalling.

4.1 In-channel Signalling

In-channel signalling employs the same physical trunk or channel to carry the call and the related control signals. This is the traditional control signalling in circuit switching networks where the control signals follow the same path as the call itself. Two forms of in-channel signalling are in use: in-band and out-of-band.

- in-band signalling - the control signals use not only the same physical path, but also the same frequency band as the data. In traditional public switched telephone networks, the same 4 kHz channel used for voice carries the control information in terms of pure tones at various frequencies. Hence, control signals can go anywhere that the data signals go. In the context of packet switching networks, packets in in-

band signalling carrying control information are multiplexed to the same connection used for data transmission. As transmission facilities are shared, control signals cannot be transmitted when data is using the channel. Besides, a faulty path hampers the transmission of both data and control signals. It also blocks all kinds of control signalling for network management functions.

- out-of-band signalling - the control signals use the same physical path as the call, but a separate frequency band is allocated for signalling purposes. In public switched telephone networks, voice signals do not use the full 4 kHz bandwidth within which a separate signalling band can be allocated to send control signals exclusively. It allows simultaneous transmission of both voice and control signals on the same line, and thus continuous supervision and control of a call. This narrow signalling band however limits the control information transfer rates and only simple control messages can be accommodated.

With in-channel signalling, control signals originated by a control processor within a switch have to travel on the data channels and be extracted by the control processor at the receiving switch. Nevertheless, in-channel signalling offers a simple solution to signalling requirements in managing public switched networks. As telecommunications networks evolve to embrace more services such as call forwarding, multiparty conferencing and toll-free calls, the amount of control information increases dramatically. A more flexible and sophisticated signalling is therefore required to provide a richer set of control messages with higher transfer rates.

4.2 Common Channel Signalling

Common channel signalling is designed to eliminate the issues caused by in-channel signalling. With common channel signalling, control signals are carried over control paths completely independent of the data channels. These signalling links are dedicated to control signals and are common to a number of data channels. The characteristics of the common channel (e.g., bandwidth) can be configured to support a rich variety of signalling functions. With a permanently available common channel, transmission of control information can overlap the sequence of events for call setup process, thus reducing the call setup time compared to in-channel methods.

With common channel signalling, the control signals are handled by a control signal processor inside the switch, which are directly transferred to another control signal processor, thus bypassing the data channels. In reality, a network that is completely controlled by common channel signalling still has to employ in-channel signalling over the last transmission links to the endpoints. It informs the network users of the status of a call by dial tone, ringback, and busy signals. An example of common channel signalling is the ITU-T Signalling System Number 7 (SS7) which defines a layered signalling network structure for the ISDN system.

There are three signalling modes of operation in common channel signalling according to the association between the path taken by a signalling message and the signalling relation to which the message refers: associated mode, non-associated mode, and quasi-associated mode.

- associated mode - the control signals of all the data channels on the same path are carried by a common channel that goes along the entire length of the path and is terminated at the same switching point. In the associated mode of SS7, the messages exchanged between two adjacent signalling points[4] are conveyed over a signalling link that directly interconnects them to each other. In this mode, only user functions located at adjacent signalling points may be accessed.

- non-associated mode - the control signals are transferred on a totally separate network with its signalling links and signal transfer points[5]. Such a signalling network carries only the control information of all the data channels, and control channels can be assigned in a more flexible manner to serve the data network.

- quasi-associated mode - a limited case of the non-associated mode, the control signals are transferred on a predetermined path through the network which is fixed at a given point in time, except for rerouting caused by failure and recovery events. In the quasi-associated mode of SS7, user functions located at any signalling point may be accessed if the corresponding routing data are present.

4.3 ATM Signalling

In an ATM environment, the traditional control signalling scheme is replaced by the concept of separated planes for the segregation of user, control and management functions in the B-ISDN protocol reference model specified by the ITU-T. This model is illustrated in Figure 5, containing a user plane to transport user information, a control plane mainly composed of signalling information, and a management plane to maintain the network and to perform operational functions. In this context, signalling is related to the dynamic activities in the exchange of control information among network entities, such as setup, maintenance, and release of ATM connections. The ultimate goal of ATM signalling is to provide separation between the call and connection control, and to enable creating multipoint, multimedia connections.

Signalling in itself is the critical issue of interoperability in interconnected ATM networks. Information exchange is required between the network users and the switching nodes, and between switching nodes. These two B-ISDN interfaces are defined as the UNI and the NNI respectively for signalling purposes.

The interpretation of the signalling mechanism in ATM networks is somewhat confusing in relation to the definitions of in-channel and common channel signalling

[4] Any point in the signalling network capable of handling control messages.

[5] Points capable of routing control messages.

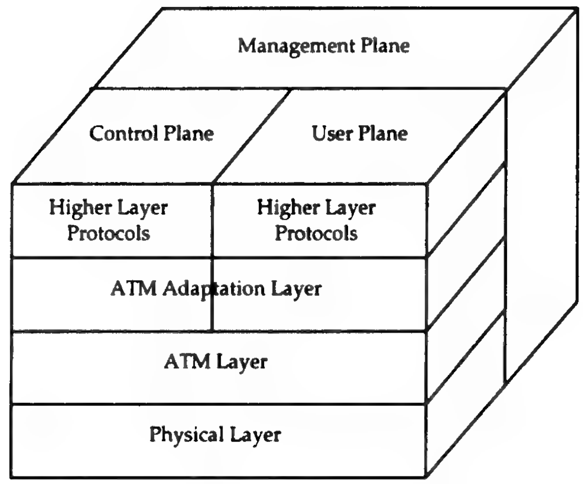


Figure 5 : B-ISDN ATM Protocol Reference Model

as stated above. Traditionally, the terms "out-of-band" signalling and "in-band" signalling have been adopted as two methods of exchanging control information in emerging packet switched networks. In-band signalling allows multiplexing packets carrying control information to the same connection for data transmission, whereas out-of-band signalling employs separate connections for control and data transmission. These two borrowed terms from circuit switched networks still remain popular in the ATM community, and thus ATM signalling is described as out-of-band in the ITU-T and the ATM Forum. Nevertheless, two types of signalling are specified in ATM networks:

- associated signalling - a signalling virtual channel (signalling VC) of each virtual path connection (VPC) is allowed to control only those virtual channels (VCs) within the same path.
- non-associated signalling - a signalling VC of each VPC controls the VCs which belong to the same VPC, as well as VCs in other VPCs.

An ATM network is expected to support both associated and non-associated signalling procedures. This is indicated in the VP-associated signalling field in a connection identifier information element (see Section 7.2) which identifies the corresponding user information flow. Associated signalling is supported only in the ITU-T Q.2931 but not in the ATM Forum UNI 3.1 or UNI Signalling 4.0 specification[6].

[6] To allow specifications to be developed and released with minimum delay, the subsequent version following UNI 3.1 is made of a number of individual components, namely, UNI Signalling 4.0, Traffic Management 4.0, and Integrated Local Management Interface 4.0 specifications.

5. Key Elements of ATM Signalling

In ATM networks, a call request to set up a connection is performed over a separate signalling virtual channel. This involves the negotiation between the user and the network with respect to the resources such as VCI/VPI, throughput, and quality of service (QoS) used for the connection. Such a common channel signalling scheme sets aside a number of reserved virtual channels (with special VPI and VCI values) for signalling, operation and maintenance, and resource management. For the UNI, the VPI is 8 bits long and the VCI 16 bits long. Within all virtual paths (VPs), the range of VCI = 0 to 15 is reserved by the ITU-T, and that of VCI = 16 to 31 is reserved by the ATM Forum as shown in Table 1.

Table 1: The range of VCI values

VCI value	reserved for
0 - 15	ITU-T defined functions
16 - 31	ATM Forum-defined functions
32 - 65535	User traffic

Unassigned or empty cells are used on physical links that have frame structures (generally PDH and SDH) by setting both VPI and VCI values to zero. VCI = 1 is reserved for the use of meta-signalling, which is the process of setting up and releasing additional signalling virtual channel connections. VCI = 2 is reserved for general broadcast signalling for the broadcasting of signalling information to all ATM users at the UNI. In both meta-signalling and general broadcast signalling, VPI = 0 designates the signalling is communicated with the local switch. The default signalling virtual channel, identified by VCI = 5 and VPI = 0, is used for all UNI point-to-point and point-to-multipoint signalling from an ATM endpoint to the local network. With any VPI value other than zero, VCI = 5 is used for point-to-point signalling with other endpoints and other networks. These reserved VPI and VCI values are summarised in Table 2.

A call setup transmits signalling messages to one or multiple destinations. This request

Table 2: Pre-defined VPI and VCI values

VPI value	VCI value	use
0	0	unassigned cells
0 (default)	1	meta-signalling with the local switch
non-zero	1	meta-signalling with other users or remote networks
0 (default)	2	general broadcast signalling with the local switch
non-zero	2	general broadcast signalling with other users or remote networks
0 (default)	5	signalling with the local switch
non-zero	5	signalling with other users or remote networks

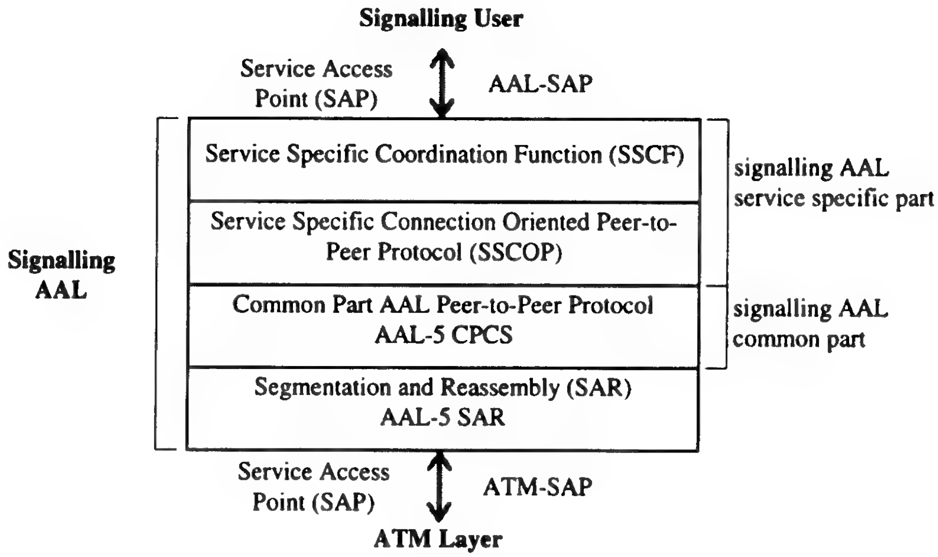


Figure 6 : Signalling AAL Structure

setup includes the selection of an appropriate VP, the mapping of VCIs and VPIs in intermediate switches, resource negotiations between switches, and the call and connection control in the calling and called ATM end systems.

6. Signalling ATM Adaptation Layer

To establish, maintain, and terminate ATM connections across a signalling interface, a signalling user (i.e., higher layer) runs on top of the signalling ATM adaptation layer (AAL) which provides reliable transport for signalling traffic. Since signalling information must not get lost or corrupted in delivery, AAL-0 or null AAL that performs no AAL function cannot be used for network control and service messaging. The signalling ATM adaptation layer provides a series of AAL functions supporting signalling connections between ATM switches, and between an ATM endpoint and an ATM switch (network). The structure of the signalling AAL is illustrated in Figure 6.

The signalling AAL defines how to transfer the signalling information reliably using ATM cells on signalling virtual channels. Similar to the AAL for individual classes of network traffic, the signalling AAL consists of a service specific part, a common part and a segmentation and reassembly (SAR) sublayer.

- The service specific part consists of the service specific coordination function (SSCF) and the service specific connection oriented protocol (SSCOP). Specified in ITU-T Recommendation Q.2130, the SSCF is a "glue" layer that connects the SSCOP protocol to the signalling user. A number of different SSCFs may be available depending on the needs of the particular signalling user. The SSCOP functions as a

peer-to-peer protocol to transfer signalling information between any pair of SSCOP entities, as specified in ITU-T Recommendation Q.2110.

- The common part is an AAL peer-to-peer protocol that assures integrity at the frame level by building header and trailer records onto the signalling user data frame.

- The segmentation and reassembly (SAR) function converts the common part convergence sublayer (CPCS) frames into cells. Cell headers and trailers are also added to provide integrity at the cell level.

The common part AAL peer-to-peer protocol and the SAR function are only trivially different from AAL-5 CPCS and SAR sublayers as specified in ITU-T Recommendation I.363, which together provide simple AAL functions for both connectionless and connection oriented data.

7. User-to-Network Signalling

ATM signalling protocols vary by the type of ATM link. ATM UNI signalling is used between an ATM end system and an ATM switch across an ATM UNI, whereas ATM NNI signalling is used across NNI links. UNI signalling is used to dynamically establish, maintain, and terminate connections between end stations and the switches at the edges of the network. Further, UNI signalling requests are carried across the UNI in a well known default connection: VPI=0, VCI=5 for both point-to-point and point-to-multipoint connections.

The offered service can be either point-to-point bidirectional or point-to-multipoint unidirectional. Point-to-point call control messages consist of signalling messages for call establishment, call clearing, and maintenance. Point-to-multipoint connections are established first by establishing a single point-to-point connection. Special messages are subsequently used to add new endpoints and to clear end parties of the connection.

7.1 Signalling Messages

Signalling information is conveyed in terms of predefined functions within a signalling message. The format of a signalling message is illustrated in Figure 7, consisting of the following information elements:

- *protocol discriminator* (1 octet) - this is to distinguish messages for B-ISDN user-network call control from the other types of messages of other protocols.

- *call reference* (4 octets) - it enables the originating end system to easily identify the call at the local user-network interface to which the particular message applies. It is therefore possible to have multiple outstanding requests at one time. Its length is

8	7	6	5	4	3	2	1	octets
Protocol discriminator								1
0	0	0	0	Length of call reference				2
Flag	Call reference value							3
Call reference value (continued)								4
Call reference value (continued)								5
Message type								6
Message type (continued)								7
Message length								8
Message length (continued)								9
Variable length information elements as required								etc.

Figure 7: General Signalling Message Format

indicated in the *length of call reference* field in octets. The *call reference value* is assigned by the originating side of the interface for a call and remains fixed and unique for the lifetime of a call. The call reference *flag* indicates which end of the signalling virtual channel allocated the call reference value.

- *message type* (2 octets) - this is to identify the function of the message being sent. Various messages are associated with point-to-point and point-to-multipoint connections such as SETUP, CALL PROCEEDING, CONNECT, ADD PARTY and RELEASE.
- *message length* (2 octets) - it specifies the length of the contents of this signalling message in octets, excluding the first nine octets occupied by *protocol discriminator, call reference, message type* and *message length*. In effect, it identifies the length of additional information elements (if any) to follow.
- *variable length information elements* - the different information elements specific to this message type are included as required.

A number of signalling messages have been specified in the ATM Forum UNI Signalling for point-to-point and point-to-multipoint call processing, as illustrated in Table 3.

7.2 Information Elements

The protocol discriminator, call reference, and the message type are mandatory information elements identifying each signalling message, as noted in Section 7.1. There are nevertheless a plethora of other information elements needed to support signalling functions. Each type of signalling messages possibly contains a number of information elements, of which some are mandatory and some are optional. A complete list of information elements is presented in the Appendix.

Table 3: UNI Signalling 4.0 Signalling Messages

Message Type	Name	Definition
Call establishment messages	SETUP	initiates call establishment
	ALERTING[7]	the called user alerting has been initiated
	CALL PROCEEDING	the requested call establishment has been initiated and no more call establishment information will be accepted
	CONNECT	call acceptance
	CONNECT ACKNOWLEDGE	user has been awarded the call
Call clearing messages	RELEASE	clears the end-to-end connection
	RELEASE COMPLETE	indicates the VC and call reference are released and the VC is available for reuse
	RESTART	requests to release all resources associated with the indicated VCs controlled by the signalling channel
	RESTART ACKNOWLEDGE	response to a RESTART message to indicate requested restart is complete
Miscellaneous messages	NOTIFY[7]	indicates information pertaining to a call/connection such as user suspended
	STATUS	response to STATUS ENQUIRY message or reports certain error conditions
	STATUS ENQUIRY	solicits a STATUS message
Point-to-Multipoint messages	ADD PARTY	adds a party to an existing connection
	PARTY ALERTING[7]	indicates the called party alerting has been initiated
	ADD PARTY ACKNOWLEDGE	response to ADD PARTY message to indicate that the add-party request was successful
	ADD PARTY REJECT	response to ADD PARTY message to indicate that the add-party request was not successful
	DROP PARTY	clears a party from an existing point-to-multipoint connection
	DROP PARTY ACKNOWLEDGE	response to DROP PARTY message to indicate that the party was dropped from the connection
	LEAF SETUP REQUEST[7]	initiates leaf joining procedures
	LEAF SETUP FAILURE[7]	indicates failure to join the leaf to the request leaf initiated join call

[7] Not supported in the ATM Forum UNI 3.1 Specification.

The format of a general information element is shown in Figure 8. Each information element is uniquely coded in the *information element identifier* field (1 octet) under the specified *coding standard*. The information element (IE) compatibility instruction is identified by the *IE instruction field* (5 bits), which determines the error handling procedures in case of unrecognised information element identifier or unrecognised information element contents. The *length of information elements* (2 octets) indicates its length excluding the first four octets of this information element.

8	7	6	5	4	3	2	1	octets
Information element identifier								1
1	Coding standard		IE instruction field					2
Length of information elements								3
Length of information elements (continued)								4
Contents of information elements								5 etc.

Figure 8: General Information Element Format

Essentially, the functionality of signalling messages is described in the attached information elements. To illustrate this, the SETUP message, one of the more important messages in UNI signalling, is described in more details. To initiate a call at the UNI, the calling user sends the SETUP message to the network which identifies the destination node and characterises the ATM connection. The SETUP message content and its constituent information elements are shown in Table 4.

The called party number IE identifies the called party. The subaddress of the called party is indicated in the called party subaddress IE, supporting both ATM endsystem address and OSI NSAP formats. This means of identification also applies to the calling party number and the calling party subaddress IEs for the calling party. The ATM connection across the UNI is identified by the connection identifier IE.

The broadband bearer capability IE indicates a requested broadband connection oriented bearer service to be provided by the network. Broadband high layer information and broadband low layer information IEs are used for compatibility checking for higher layer information type and the layer 2 protocol respectively.

The requested AAL and its parameters for the ATM connection are indicated in the AAL parameters IE, including the forward and backward maximum CPCS-SDU size (for AAL-3/4 and AAL-5), the value of MID[8] range (for AAL-3/4), and user-defined AAL information. The contents of this IE are transparent to the network, except for the case of interworking.

[8] The multiplexing identification (MID) field in the AAL 3/4 SAR-protocol data unit (SAR-PDU) structure is 10 bits long and is used to allow multiplexing AAL connections into a single ATM connection.

Table 4 : Information Elements of SETUP Message (UNI Signalling 4.0)

Information Element	*Direction*	*Type*	*Length*
Protocol discriminator	Both	M	1
Call reference	Both	M	4
Message type	Both	M	2
Message length	Both	M	2
AAL parameters	Both	O(1)	4-20
ATM traffic descriptor	Both	M	12-30
Broadband bearer capability	Both	M	6-7
Broadband high-layer information	Both	O(2)	4-13
Broadband repeat indicator	Both	O(3)	4-5
Broadband low-layer information	Both	O(4)	4-17
Called party number	Both	M	(5)
Called party subaddress	Both	O(6)	4-25
Calling party number	Both	O(7)	4-26
Calling party subaddress	Both	O(8)	4-25
Connection identifier	$n \rightarrow u$	M	9
End-to-End transit delay*	Both	O(9)	4-11
Extended QoS parameters*	Both	O(10)	4-25
QoS parameter*	Both	O(11)	4-6
Broadband sending complete	Both	O(12)	4-5
Transit network selection	$u \rightarrow n$	O(13)	4-9
End point reference	Both	O(14)	4-7
Narrowband bearer capability*	Both	O(15)	4-14
Narrowband high layer compatibility*	Both	O(16)	4-7
Narrowband low layer compatibility*	Both	O(17)	4-20
Notification indicator*	Both	O(18)	4-6
Progress indicator*	Both	O(19)	4-6
Generic identifier transport*	Both	O(20)	4-33

* Not supported in the ATM Forum UNI 3.1 Specification.

Note 1 - Included in the user-to-network direction when the calling user wants to pass ATM adaptation layer parameters information to the called user. Included in the network-to-user direction if the calling user included an ATM adaptation layer parameters IE in the SETUP message.

Note 2 - Included in the user-to-network direction when the calling user wants to pass broadband high layer information to the called user. Included in the network-to-user direction if the calling user included a broadband high layer information IE in the SETUP message.

Note 3 - Included when two or more broadband low layer information IEs are included for broadband low layer information negotiation.

Note 4 - Included in the user-to-network direction when the calling user wants to pass broadband low layer information to the called user. Included in the network-to-user direction if the calling user included a broadband low layer information IE in the SETUP message. Two or three IEs may be included in descending order of priority, i.e., highest priority first, if the broadband low layer information negotiation procedures are used.

Note 5 - Minimum length depends on the addressing. Maximum length is 25 octets.

Table 4 : Information Elements of SETUP Message (UNI Signalling 4.0) (continued)

Note 6 - Included in the user-to-network direction when the calling user wants to indicate the called party subaddress. Included in the network-to-user direction if the calling user included a called party subaddress IE in the SETUP message. If the called party subaddress IE is used to convey an ATM endsystem address, then an additional called party subaddress IE may be present to convey an OSI NSAP or user specified subaddress.

Note 7 - May be included by the calling user or by the network to identify the calling user.

Note 8 - Included in the user-to-network direction when the calling user wants to indicate the calling party subaddress. Included in the network-to-user direction if the calling user included a calling party subaddress IE in the SETUP message. If the calling party subaddress IE is used to convey an ATM endsystem address, then an additional calling party subaddress IE may be present to convey an OSI NSAP or user specified subaddress.

Note 9 - Included to specify an end-to-end transit delay requirement.

Note 10 - Included to specify individual QoS parameter requirements for the call. This IE is mandatory when the QoS parameter IE is absent.

Note 11 - This IE is mandatory when the extended QoS parameter IE is absent. When the extended QoS parameter IE is present, the QoS parameter IE may be included to facilitate interworking with networks that do not support the extended QoS parameter IE.

Note 12 - It is optional for the user to include the broadband sending complete IE when en bloc sending procedures (i.e., complete address information is included) are used; its interpretation by the network is optional. It is optional for the network to include the broadband sending complete IE when en bloc receiving procedures (i.e., complete address information is included) are used.

Note 13 - Included by the calling user to select a particular transit network.

Note 14 - Not used for point-to-point connection establishment. Must be included in SETUP messages involved in point-to-multipoint connection establishment.

Note 15 - Mandatory for N-ISDN services. May be repeated if the narrowband bearer capability negotiation procedure is used.

Note 16 - Included in the user-to-network direction when the calling user wants to pass narrowband high layer compatibility information to the called user. Included in the network-to-user direction if the calling user included a narrowband high layer compatibility IE in the SETUP message.

Note 17 - Included in the user-to-network direction when the calling user wants to pass narrowband low layer compatibility information to the called user. Included in the network-to-user direction if the calling user included a narrowband low layer compatibility IE in the SETUP message.

Note 18 - This indicator may be present whenever notification is delivered.

Note 19 - Included in the event of interworking or in connection with the provision of in-band information or patterns.

Note 20 - Included in the user-to-network direction when the calling user wants to pass generic identifier transport information to the called user. Included in the network-to-user direction if the calling user included a generic identifier transport IE in the SETUP message.

The forward and backward QoS class is specified in the QoS parameters IE, indicating unspecified class or B-ISDN service classes[9] A, B, C, and D for each direction of the connection. The maximum acceptable end-to-end transit delay is described in the end-to-end transit delay IE.

The ATM traffic descriptor IE indicates a set of traffic parameters for each direction. These include the forward and backward peak cell rates (PCRs)[10], sustainable cell rates (SCRs)[11], maximum burst size, and a traffic control capability such as the best effort service indicator and the use of tagging the CLP bit at the ATM cell header. The connection characteristics can be negotiated by including either the minimum acceptable ATM traffic descriptor IE or the alternative ATM traffic descriptor IE. If the network is not able to provide the traffic parameter values specified in the ATM traffic descriptor IE, the lowest values that the user is willing to accept for the call/connection are specified in the minimum acceptable ATM traffic descriptor IE for traffic parameter negotiation. On the other hand, an alternative ATM traffic descriptor is specified in the alternative ATM traffic descriptor IE for additional consideration of traffic parameters during call/connection setup.

The establishment of available bit rate (ABR) call/connection is specified in the ABR setup parameters IE and ABR additional parameters IE for additional negotiation. The ABR service offers a guaranteed delivery service (with minimal cell loss) to users who can tolerate a widely varying throughput rate. This is the available bandwidth in the running network after other traffic utilising guaranteed bandwidth services has been serviced.

7.3 Basic Point-to-Point Call

Signalling messages provide the network with enough information to characterise the source and to locate the destination UNI. The functions provided in ATM signalling for call and connection control cover call setup, call clearing and status. We now consider the signalling procedure for the point-to-point channel connection. These three groups of point-to-point signalling messages are given in Table 3. Each message contains a number of mandatory or optional information elements describing various aspects of the interaction.

Consider the sequence of events and the messages exchanged to establish a point-to-point connection between two end systems across a network, as illustrated in Figure 9. End systems 1 and 2 are connected to the ATM network across two separate UNIs. End

[9] According to time relation between source and destination, bit rate, and connection mode, the ITU-T has defined four generic classes of network traffic: Class A for constant bit rate circuit emulation; Class B for variable bit rate voice, video, and multimedia; Class C for connection oriented data; and class D for connectionless data services.

[10] The PCR of a connection is the inverse of the minimum time between two cell submissions to the network.

[11] The SCR is an upper bound on the conforming average rate of an ATM connection, over time scales that are long relative to those for which the PCR is defined.

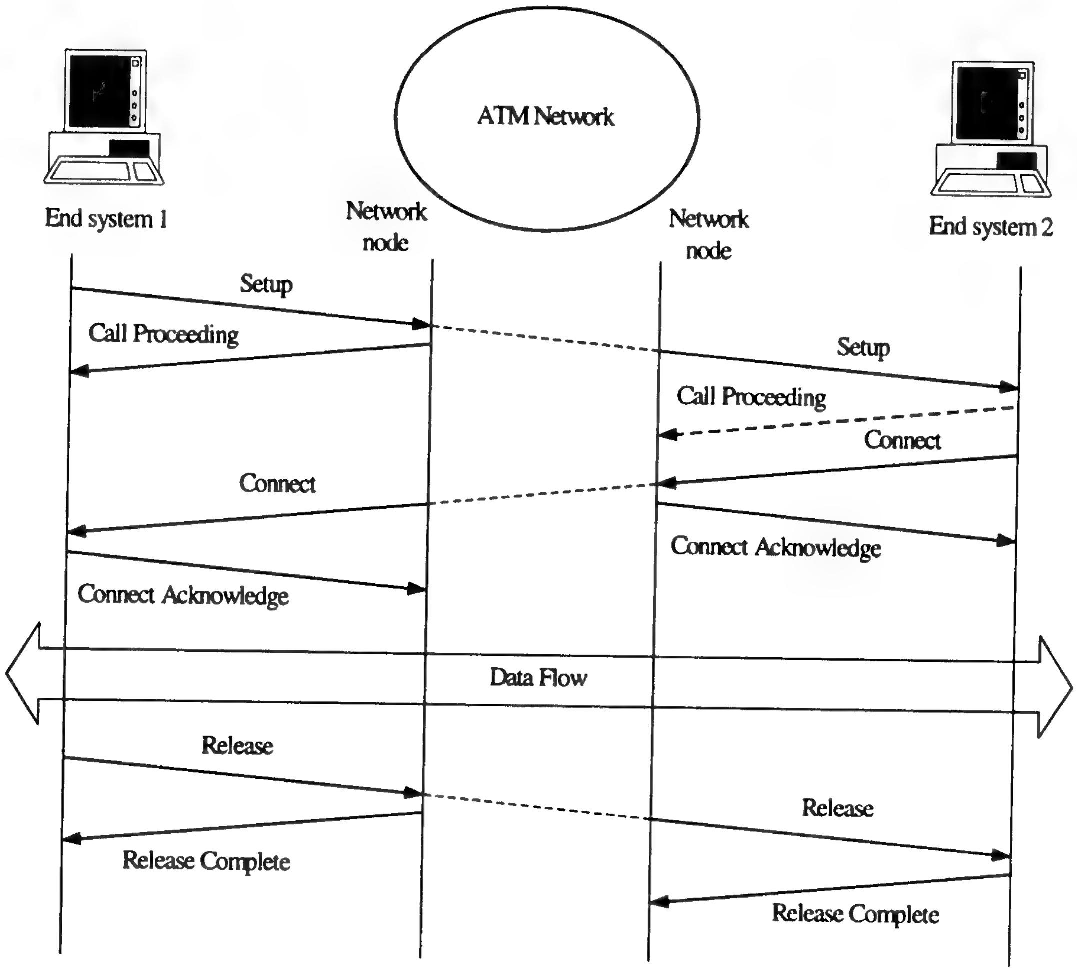


Figure 9: Signalling Messages for Point-to-Point ATM Switched Virtual Circuits

system 1 initially requests a call by sending a SETUP message across its UNI to the network, which contains pertinent information that identifies the two end nodes and connection characteristics. The network node replies to end system 1 with a CALL PROCEEDING message with the VPI/VCI value to be used for the user traffic. The destination network node is informed and a SETUP message is formulated and delivered to end system 2 across its UNI. This SETUP message includes the VPI/VCI value used for data transmission. If necessary, end system 2 sends to the network a CALL PROCEEDING or ALERTING[7] message if it will take longer to process.

At this time, the call setup can be rejected by the network or end system 2. In case the call cannot be accepted by the network, the originating network node sends a RELEASE COMPLETE message to end system 1 across its UNI. If end system 2 is unable to connect, a RELEASE COMPLETE message is sent by end system 2 across its UNI to indicate that the call setup is rejected. The network then conveys the rejection

information to the originating network node where a RELEASE COMPLETE message is sent across the UNI to end system 1 to report a call rejection.

For a successful connection setup, end system 2 forwards a CONNECT message accepting the setup request. In response, the destination network node sends a CONNECT ACKNOWLEDGE message to end system 2. Using the designated VPI/VCI value, end system 2 can start sending data on the ATM connection. The acceptance information is conveyed back to the originating network node, and a CONNECT message is sent to end system 1 containing all information related to the connection setup.

End system 1 accepts the connection setup by sending a CONNECT ACKNOWLEDGE message to the network node. End system 1 can start sending data on the ATM connection using the assigned VPI/VCI value in the CALL PROCEEDING or CONNECT message. Data cells can now flow across the UNIs in both directions.

Any endpoint (say, end system 1) who wishes to terminate the connection sends a RELEASE message across its UNI. The network node then sends a RELEASE COMPLETE message and clears the connection. A RELEASE message is also sent to end system 2 across its UNI. End system 2 sends a RELEASE COMPLETE message and clears the connection accordingly.

7.4 Point-to-Multipoint Call

A point-to-multipoint connection contains a number of unidirectional VC or VP links with associated end nodes. The calling party, also known as the root, generates traffic to all other end nodes, also referred to as leaf nodes. Leaf nodes cannot communicate with the root or with each other directly through the point-to-multipoint connection.

7.4.1 Root as Initiator

Leaves join a point-to-multipoint connection under the initiation of the root. A point-to-multipoint connection is set up by first establishing a point-to-point connection from the root to a leaf node. This is basically identical to setting up a point-to-point connection. The SETUP message is constructed with endpoint reference set to zero, and broadband capability information element indicating a point-to-multipoint connection in its user plane connection configuration field. Once a point-to-point connection is established, additional leave can be added using ADD PARTY messages. The same VPI/VCI values are used at the UNI to reach all leaf nodes, therefore all parties are sharing the same virtual connection, with the same connection identifier, QoS, bearer capability, and ATM traffic descriptor. As such, each ADD PARTY message has the same call reference value as in the original SETUP message. However, an ADD PARTY message can only be used on an active connection, and a SETUP message is still required to allocate new resources between switches and across a user-network interface [3].

If the network rejects an add party request, an ADD PARTY REJECT message is sent to the root. If the requested service can be provided by the network to the new leaf node, an ADD PARTY message initiated from the root is delivered as an ADD PARTY message or a SETUP message to the called party across its user-network interface. The called party accepts the point-to-multipoint connection by sending an ADD PARTY ACKNOWLEDGE message or a CONNECT message to its network node. Across the network, the root is notified by an ADD PARTY ACKNOWLEDGE message from the network side of the UNI. In case the called party rejects the connection request, an ADD PARTY REJECT message or a RELEASE COMPLETE message is sent by the called party to the network. Accordingly, the root receives an ADD PARTY REJECT message across its UNI.

The root releases a particular leaf node by sending a DROP PARTY message with the corresponding endpoint reference to the network. In response to this, the network returns a DROP PARTY ACKNOWLEDGE message to the root. This DROP PARTY message propagates across the network to clear the called party, and a DROP PARTY ACKNOWLEDGE message is returned back to the network. Finally, the root clears the call by sending a RELEASE message to the network, and the network side of its UNI disconnects the virtual channel and returns a RELEASE COMPLETE message as in a point-to-point connection.

7.4.2 Leaf Initiated Join (LIJ) Capability

The ATM Forum UNI Signalling 4.0 provides a capability where leaves can join point-to-multipoint connections with or without intervention from the root of the connection. The leaf initiated join capability allows two modes of operations:

- leaf-prompted join without root notification - This type of connection is referred to as a Network LIJ connection. A leaf initiates a request over its UNI to join a point-to-multipoint connection. A request to join an existing connection is handled only by the network, not the root. A request to join a new connection however involves the initial set up by the root of the connection.
- root-prompted join - Referred to as a Root LIJ connection, this type of connection involves the supervision by the root. A leaf submits a request to join a point-to-multipoint connection over the UNI. Normal point-to-multipoint procedures are initiated by the root to add leaves to or remove leaves from a new or established connection.

The root can either create a point-to-multipoint call as a Network LIJ call, which supports network joining of leaves, or as a Root LIJ call, where all subsequent LEAF SETUP REQUEST messages will be forwarded to the root. A Network LIJ call is specified in the LIJ parameters information element which is included in the initial SETUP message originated from the root, otherwise the creation of a Root LIJ call is assumed.

To join a point-to-multipoint call, a leaf sends a LEAF SETUP REQUEST message using the dummy call reference[12] across the leaf interface. If a leaf attempts to join a Root LIJ call or a new Network LIJ call, the LEAF SETUP REQUEST message is forwarded by the network to the root. If the leaf setup request is considered successful, a SETUP or ADD PARTY message is transmitted to the leaf to add it to the requested LIJ call. If the network or the root rejects the join request for any reason, a LEAF SETUP FAILURE message is returned to the leaf using the dummy call reference.

8. Network Node Signalling in Private Networks

The network node signalling is the process that controls the signalling information among network nodes. These network nodes are coordinated to work together to establish, modify, and release connections consistently and efficiently. The need to standardise well-defined interfaces is apparent. Consider a situation where a foreign device is connected to the other side of an network node interface (NNI), the network must be protected from any harmful operations initiated by the device that might disrupt the network operation or compromise its performance. Private and public NNIs are two defined interfaces distinct from one another in this sense.

Signalling in public networks is defined between two networks, though it is often used internally in a public network as well. Both types of interfaces are referred to as public NNIs. The corresponding interface between two private networks is a private NNI (PNNI) defined by the ATM Forum. A PNNI is meant to be a private interface such that the connected switches are administered by the same user organisation, whereas a public NNI isolates a foreign device from the internals of the network. The private NNI protocol is being specified by the ATM Forum and the public NNI is being specified by the ITU-T. One of the major differences is that in the case of the public NNI, there is going to be a strong dependence on the signalling system number 7 (SS7).

The ATM Forum version of public NNI is a carrier-to-carrier interface called the broadband inter-carrier interface (B-ICI). This specification includes the various physical layer interfaces, ATM layer management, and higher layer functions required for the interworking between ATM and various services.

8.1 Private Network-to-Network Interface (PNNI)

Although each network may use its own proprietary signalling to establish connections internally among its ATM switches, the standardisation of the switch-to-switch interface would ensure the interoperation between switches and therefore private networks. For this purpose, the private network-network interface (PNNI[13]) is

[12] Coded with all bits of the call reference value set to 1.

[13] PNNI stands for either Private Network Node Interface or Private Network-to-Network Interface, reflecting these two possible usages.

an ATM Forum specification for connecting either ATM nodes (switches) or ATM networks. Thus, the PNNI can be used either as an interface between different ATM networks or as an interface between ATM switches within a single ATM network. Implementing this interface enables the construction of scalable full-function networks of arbitrary size and complexity. Including both routing and signalling frameworks, the ATM Forum PNNI phase 1 specification is designed to be compatible with UNI 3.1 specification (but not UNI Signalling 4.0 at this release).

The PNNI routing protocol is used to route various network control information, such as that related to call admission, path selection, and topology distribution through the ATM network. The PNNI signalling protocol is the other component of the PNNI specification which is used to relay topology and routing information, as well as call control functions for connection establishment, management, and termination, between PNNI nodes within the networks.

8.1.1 PNNI Routing

The PNNI routing protocol is used to distribute topology and link state information of the network, enabling each node to have sufficient topology information to process a call setup request. The PNNI routing protocol is developed to allow for high scalability and to support QoS-based routing. The PNNI routing protocol determines the route on which an ATM connection is to be set up, and along which the data will flow. The connection request is routed along a path that leads to the destination, while supporting the QoS requested in the connection setup.

Every PNNI switch carries a topology database and network directory sufficient in scope to do some basic route planning. Detailed planning is a cooperative process in which many switches along the route participate. Every switch maintains a database which reflects a partial view of the network. The route is made more precise as the connection setup request is forwarded and processed by the switches along the chosen route. The switches along the route add detail to the route consistent with their knowledge of the topology.

8.1.1.1 PNNI Network Topology

The PNNI routing hierarchy is designed to provide for efficient routing, and reduce the overhead of maintaining the entire topology of the network at each node. This includes information for every physical link and reachability information for every node in the network. The PNNI routing protocol allows scalability by summarising reachability information between levels in the network hierarchy. The QoS and reachability information is flooded only to all members of a peer group in terms of PNNI topology state packets (PTSPs).

In Figure 10, an ATM network consisting of 18 interconnected physical switches, or lowest-level nodes is shown in three different views. Basically, the 19 most significant octets of ATM end system addresses (AESAs) identify the end systems attached to

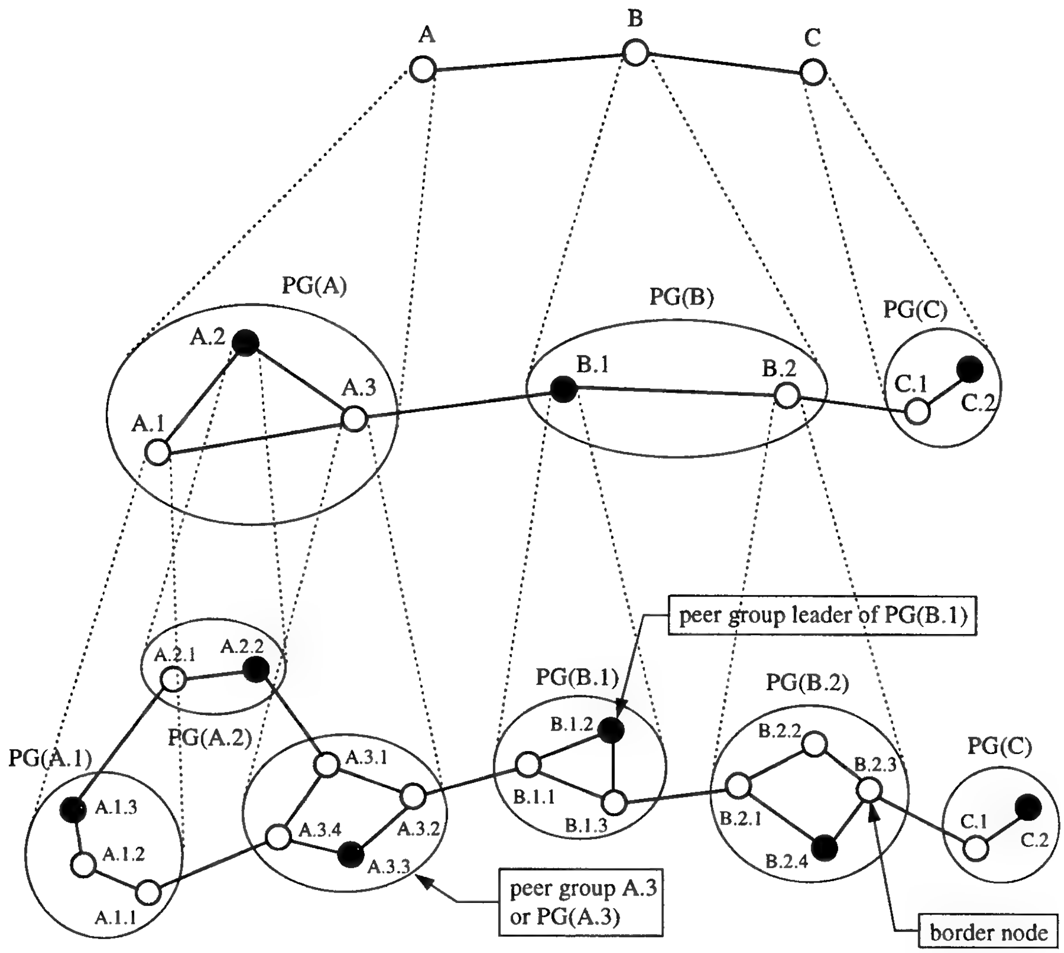


Figure 10 : Views of Hierarchical Network

physical switches, and hence determine the PNNI route across the network. This is the real network topology at the physical level which is simple but flat. Each lowest-level node has to maintain the entire topology of the network, including information for every physical link and reachability information for every node. Inevitably, enormous overhead is created for large networks. The PNNI routing hierarchy is designed to reduce this overhead while providing for efficient routing.

This physical level, or the lowest level, of the network can be structured hierarchically by organising the physical switches into peer groups A.1, A.2, A.3, B.1, B.2, and C. Each node exchanges information with other members of the group such that all members maintain an identical view of the group. A peer group leader is elected in each peer group to represent the entire peer group in the next hierarchical level. Such an arrangement forms the next level of peer group hierarchy in which the network is

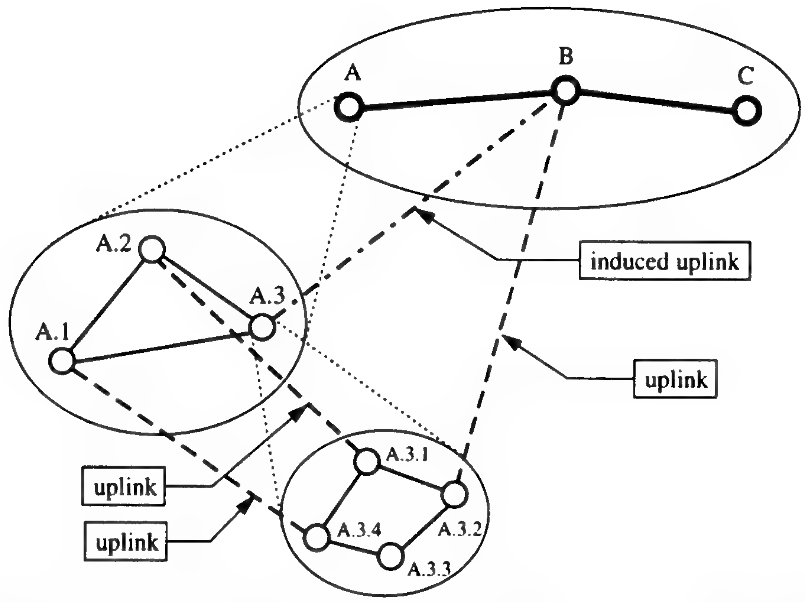


Figure 11 : View of the Global Topology from Nodes A.3.1, A.3.2, A.3.3, and A.3.4

composed of seven logical nodes[14]. As an abstraction of a peer group, a logical group node represents that peer group in the next PNNI routing hierarchy level. For example, logical group node A.2 represents peer group A.2 in the next higher level. Logical group node A.1, A.2, and A.3 are also organised into peer group A. The functions of the logical group node and the peer group leader of its child peer group are in fact executed in the same switching system. A logical group node floods the aggregated and summarised information about its child peer group into its own peer group. Conversely, a logical group node passes information received from its peer group to the peer group leader of its child peer group for flooding. At the highest level, the entire network is encompassed in a single peer group consisting of only three logical group nodes A, B, and C. In this example, each level is itself a valid view of the network such that the total network is constructed as a hierarchy of logical group nodes.

8.1.1.2 Single Node Perspective

The PNNI routing hierarchy enables topology aggregation and therefore significant reduction of topology database at each node and its update traffic. This results in a restricted view of the total network topology in each switch. Figure 11 shows the view corresponding to the network description contained in the topology database of each member in the lowest peer group A.3. Each of these switches gathers an exact topology only for its own peer group but a summarised topology for the remainder of the network. In general, the perspective that a peer group has of the rest of the network

[14] As a convenient way of representation, there is no middle level in this example hierarchy making up logical node C which only consists of two physical switches. It is not necessary to have a uniform layering structure through the network.

corresponds to all ancestor peer groups of that group. In this example, these include peer group A and the highest level peer group. In addition, the links A.3.4 to A.1, A.3.1 to A.2, and A.3.2 to B are considered uplinks which represent logical connections to higher-level nodes. The link A.3 to B is derived from uplink A.3.2 to B and therefore called an induced uplink.

8.1.1.3 Exchange of PNNI Routing Information

Summarised topology information is exchanged between logical group nodes within a peer group over routing control channels (RCCs). At the lowest level of the PNNI routing hierarchy, neighbouring nodes or physical switches are directly connected either by a physical link or a virtual path connection[15]. A permanent VC connection is used as the routing control channel for the exchange of PNNI routing protocol packets between logically or physically adjacent nodes. A reserved VC connection with VPI=0 and VCI=18 is used over physical links, whereas the exchange of the PNNI routing protocol over a VP connection with , for example, VPI=V, takes place over the PNNI VCC within the VPC, that is VPI=V and VCI=18. At higher levels, routing control messages are exchanged between logical group nodes by setting up an established switched VC connection with the VPI and VCI assigned by signalling. These routing control channels are switched through other nodes at lower levels.

A Hello protocol runs continuously between neighbouring nodes at each level over the routing control channels. PNNI Hello packets are used to discover and verify the identity of neighbouring nodes and to determine the status of the links to those nodes. Most importantly, when a node receives a Hello packet from its neighbour, the peer group identifier (PGID) received from its neighbour is compared with its own PGID to determine whether this adjacent node is part of its peer group or not. Accordingly, the Hello protocol enables border nodes to discover logical group nodes at higher levels. These Hello packets are sent at regular intervals to ensure that an accurate picture of the network state is maintained.

A topology state routing protocol informs all nodes of the knowledge about reachability and resource availability within the network. Each node bundles its local QoS and reachability information in PNNI topology state elements (PTSEs) which are encapsulated within PNNI topology state packets (PTSPs) for flooding throughout its peer group. PTSPs are transmitted at regular intervals, or when triggered by particular events. A collection of all PTSEs received at a node constitutes its topology database, indicating the current state of the global network from such a node. This information is sufficient to compute a route from the given node to any other attached node in the PNNI routing domain.

8.1.1.4 Path Selection

---

[15] It is possible to have a VP connection as a logical link between switches instead of a direct physical one.

An ATM network is expected to guarantee QoS parameter values of the established calls. When setting up a new call, the path is chosen to satisfy the QoS requirements as well as to meet network efficiency criteria, according to the current available information. PNNI uses source routing for all connection setup requests, in which the source switch is responsible for selecting the path to the destination, on the basis of its local knowledge of the network topology gained by the PTSPs.

In a switch, the function that determines whether there is sufficient local capacity for the new connection is called connection admission control (CAC). During connection setup, each switch along the chosen path performs CAC to ensure that accommodating the connection does not violate the QoS guarantees to the established connections. Accepting a connection also changes the ability of a switch to accept future connections. This event will trigger origination of new PTSE instances, informing others of the updated resource availability of this node.

Based on the current knowledge of a source switch, the generic connection admission control (GCAC) algorithm determines whether a link or node is likely to have potentially enough resources to support the proposed connection. It calculates the expected CAC behaviour of another node, so as to determine a possible path to the destination which can provide the requested QoS. If such a path is found, a designated transit list (DTL) that describes the complete route is inserted into the signalling request and forwarded by the node along the path. Each node in the path still performs its own CAC on the routed request because its own state may have changed since it last advertised. If an immediate node along the path is unable to follow the DTL due to the lack of resources, the selected path is rolled back to the earlier node in the path, which then attempts to discover another path to the destination. This mechanism is called crankback.

8.1.1.5 PNNI Addresses and Identifiers

PNNI operates in a topologically hierarchical environment. The key to constructing the PNNI routing hierarchy is in the PNNI addressing and identification based on the ATM end system addresses (AESAs). When address assignment follows the PNNI topological hierarchy, it is possible to allow an address prefix to represent reachability to all addresses that begin with the stated prefix. Shorter prefixes of AESAs summarise greater numbers of addresses and vice versa.

ATM private addresses are used to uniquely identify various physical and logical resources including ATM end stations, switches and logical groupings of switches. PNNI routing only operates on the first 19 octets of 20-octet AESAs. The last octet is the selector which has only local significance to the end system and is ignored by PNNI routing. A unique AESA is required to address any individual node in the PNNI hierarchy, including physical switches (the lowest level nodes) and logical group nodes. If a physical switch simultaneously acts as a peer group leader within multiple levels of peer groups, a unique ATM address must be assigned to that node at each of these levels. For further information on ATM addressing, see [4].

PNNI entities (nodes, links, and peer groups) exist at various hierarchical levels. The structure of the hierarchy is determined by the way in which peer group identifiers (PGIDs) are associated with logical group nodes. A peer group identifier is defined as a prefix of at most 13 octets on an AESA. A level indicator specifies the exact number of significant bits in the prefix used for the peer group identifier, indicating the relative level of PNNI routing. Peer group identifiers are encoded using 14 octets: 1 octet level indicator followed by 13 octets (104 bits) of identifier information as shown in Figure 12. The value of the level indicator is within the range of zero and 104. It specifies the number of significant bits within the identifier information field that are used for the peer group identifier, while the rest of the right-most bits are all set to zero.

level (1 octet)	identifier information (13 octets)

Figure 12 : Peer Group Identifier

A PNNI node represents either a physical switch or a peer group, which is specified by a 22-octet node identifier. To represent a single physical switch, the node identifier consists of a 1-octet level indicator specifying the level of the containing peer group, a second octet of the value 160[16], and the 20-octet AESA of the system denoted by the node. For a logical group node representing a child peer group, the node identifier is constructed by a 1-octet level indicator specifying the level of the logical group node, the 14-octet peer group identifier of the child peer group, followed by the 6-octet end system identifier (ESI) of the physical system implementing the logical group node functionality, and the last octet of the value zero.

Representing the connectivity between two logical nodes, a logical link in the network is specified by the node identifier of its transmitting node and the logical port identifier. A logical port identifier is 4 octets length and has local significance only.

8.1.1.6 PNNI Hierarchy Configuration

In the PNNI network, ATM switches are logically placed in hierarchical positions. The key to constructing a PNNI hierarchy is in the ATM address. During the initial network design, the address hierarchy of the switches has to be carefully assigned for scalability to accommodate any possible growth over time. Each physical switch is uniquely identified by its ATM address. The network uses the information in the 13-octet address prefix to create the hierarchy, which uniquely identifies a physical switch and all its connected end devices. In the PNNI hierarchy, each lowest level node (i.e., physical switch) must know its own ATM address, zero or more address summaries, its node identifier, and the peer group identifier. Host reachability information is advertised by each node to all attached end systems. To allow network scalability,

[16] This recognises this case from the one below in which a peer group identifier cannot start with the value 160.

reachability to multiple hosts should be summarised in a small set of address prefixes as much as possible.

At 1 bit per level, the network planner can partition the 13-octet address prefix into any number of hierarchical levels up to 104. Each lowest level peer group may be configured, where appropriate, to have any level between 0 and 104. The default level for a node is 96. In small networks, establishing a hierarchical structure may not be necessary for network operation. However, in larger networks, ATM switches and devices should be connected according to an address hierarchy. In practice, most networks work well with only three to five hierarchical levels.

Once a level is configured in each node, a peer group identifier (PGID) can be created in default from the 1-octet level indicator and a prefix of level bits on the ATM address of the node padded with trailing zero bits. Configuration at the lowest level can be simplified if all nodes in a peer group have been assigned addresses that start with the same 12 octet prefix (the default level), which is unique to that peer group.

For nodes that are not capable of being peer group leader, the absolute minimum configuration information comprises simply the ATM address and the hierarchical level. Since a peer group leader is selected based on a priority system, each node capable of being peer group leader must know the peer group identifier of its parent peer group, and its leadership priority.

It is likely that a capable physical switch takes up the roles of peer group leaders (PGLs) at consecutive levels of the hierarchy. If such a node fails, it will affect a large number of peer groups in which that switch participates. To alleviate this avalanche of network failure, another physical switch that exists at a higher level should be given high PGL priority. Such a node can take over the duties of PGL from that level upwards.

8.1.2 PNNI Signalling

Based on the ATM Forum UNI specification, the PNNI signalling protocol contains the procedures to dynamically establish, maintain and clear ATM connections at the PNNI between two ATM networks or two ATM network nodes. The procedures are defined in terms of messages and information elements used to characterise the ATM connection.

PNNI signalling is based on the network side procedures of UNI signalling. It is therefore very similar to the signalling used at the UNI as specified by the ITU-T in Q.2931. The UNI signalling request is mapped into PNNI signalling at the source (ingress) switch, which is later remapped back into UNI signalling at the destination (egress) switch. The PNNI signalling procedure is essentially symmetric, because any PNNI node performs the same set of functions in receipt of a signalling message. In contrast, the signalling procedures are asymmetric at the UNI. Upon receiving a signalling message (for example, a SETUP message), the UNI procedures performed

differ depending on the recipient: user or network [5]. Not surprisingly, there are significant similarities between UNI and PNNI signalling messages. In fact, the set of PNNI signalling messages contains all those in the UNI signalling as mentioned in Section 7.1, except CONNECTION ACKNOWLEDGE, LEAF SETUP REQUEST, and LEAF SETUP FAILURE. The ATM Forum PNNI 1.0 specification is only compatible with UNI 3.1 specification, it therefore does not support UNI Signalling 4.0 specific signalling features such as proxy signalling, leaf initiated join capability or user to user supplementary service.

A large portion of the PNNI information elements are the same as UNI information elements (see the Appendix). There are nevertheless a few information elements specific to the PNNI:

1. called party soft PVPC or PVCC[17] (max 11 octets) - this element indicates the VPI or VPI/VCI values of a PVC segment between the called connecting point and the user of a PVPC or PVCC respectively.
2. calling party soft PVPC or PVCC (max 10 octets) - it indicates the VPI or VPI/VCI values used for the PVC segment by the calling connecting point.
3. crankback (max 72 octets) - this is to indicate that crankback procedures have been initiated, also indicates the node or link where the call/connection or party cannot be accepted, and the level of PNNI hierarchy at which crankback is being carried out.
4. designated transit list (max 546 octets) - this element indicates the logical nodes and logical links that a connection is to traverse through a peer group at some level of hierarchy.

Figure 13 illustrates the sequence of PNNI signalling messages exchanged between end system 1 and end system 2 for a successful end-to-end connection across three networks or switches. Call establishment is initiated when end system 1 sends a SETUP message to its attached network A across UNI. Network A returns to the calling user a CALL PROCEEDING or ALERTING message if it takes longer to process. If resources are available, network A selects an outgoing link while a SETUP message is sent to network B across the PNNI. This procedure continues until the SETUP message is forwarded to network C to which the called user is attached. A SETUP message is sent to end system 2 across its UNI. End system 2 may send back a CALL PROCEEDING or ALERTING message indicating that the requested call establishment has been initiated.

End system 2 accepts the call by forwarding a CONNECT message across its UNI to network C which returns a CONNECT ACKNOWLEDGE message. This procedure propagates across the individual PNNIs through the networks to network A to which the calling user is attached. A CONNECT message is delivered to end system 1 across

[17] A soft PVPC or PVCC is a permanent virtual path connection or virtual channel connection, where the establishment within the network is done by signalling.

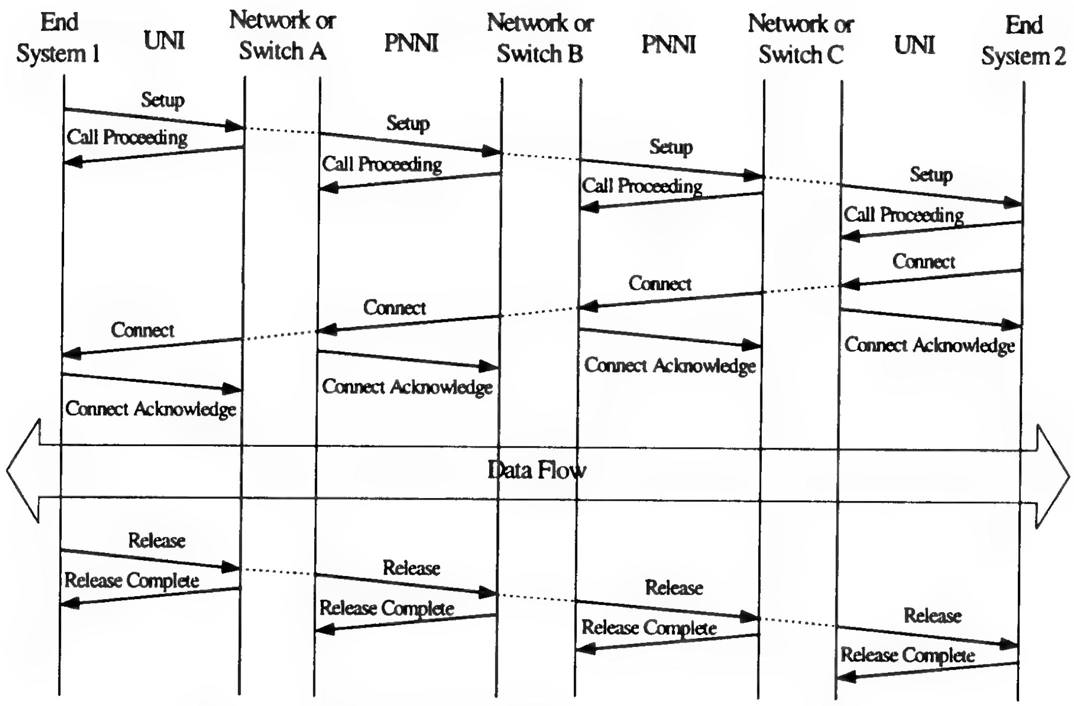


Figure 13 : Call Establishment and Release Signalling at the PNNI

its UNI. Accepting the connection setup, end system 1 returns a CONNECT ACKNOWLEDGE message and can start sending data along the end-to-end path.

The established connection is actively used until one of the end users (say, end system 1) issues a RELEASE message across its UNI. A RELEASE COMPLETE message is returned indicating that the connection is cleared. The RELEASE message is forwarded across the PNNI to the next network which returns a RELEASE COMPLETE message. Upon receiving the RELEASE COMPLETE message, a network releases all reserved network resources for the corresponding connection. This procedure repeats until network C, and a RELEASE message is delivered to end system 2 across the UNI. A RELEASE COMPLETE message is returned by the called user indicating that the resources are released and available for reuse.

8.2 Interim Inter-switch Signalling Protocol (IISP)

Also known as the PNNI Phase 0, the interim inter-switch signalling protocol (IISP) is an extension of UNI signalling. It is essentially a signalling protocol for inter-switch communication which incorporates additional information elements for such NNI-related parameters as DTL. Signalling is carried across NNI links on the same virtual channels, VCI=5, which is used for signalling across the UNI.

Signalling requests are routed between switches using address prefix tables within each switch, which are configured with the address prefixes that are reachable through

each port on the switch. On receipt of a signalling request, the switch checks the destination address against the prefix table and selects the port with the longest prefix match. The signalling request is then forwarded across that port using UNI procedures. The IISP however does not support QoS-based routing, nor crankback functionality [6].

9. Network Node Signalling in Public Networks

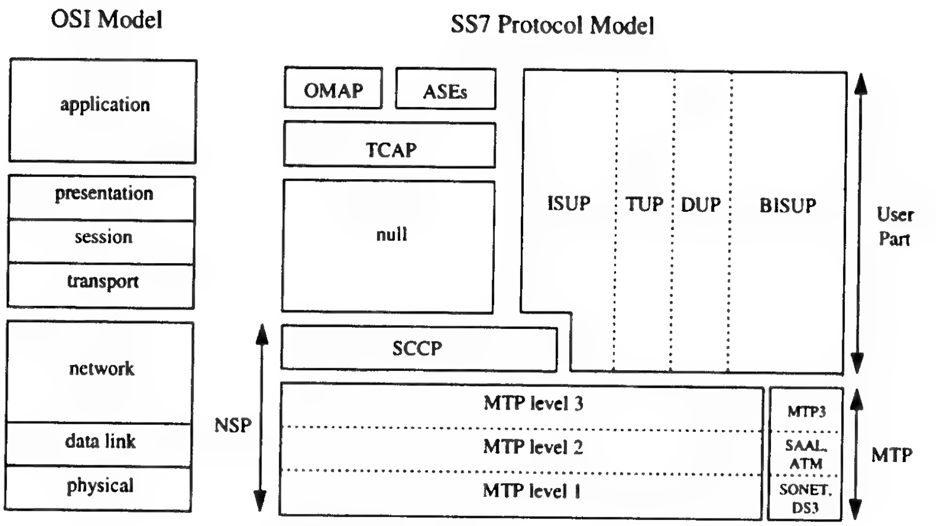
In public telecommunications networks, a special data communication architecture known as common channel signalling (CCS) is developed for the transfer of signalling information between control processors over control paths. To communicate using CCS, each one of these processing nodes requires the implementation of the necessary features of ITU-T signalling system number 7 (SS7). These are the mechanisms and protocols needed between public network switches to support end-to-end transport of the ISDN or B-ISDN signalling information received by the originating public switch, communicated from the user over the UNI. The underlying Network Node Interface (NNI) is defined by the ITU-T which only includes the SONET/SDH physical layer and the ATM layer.

9.1 SS7 Protocol Architecture

The structure of the signalling system number 7 (SS7) protocol is shown in Figure 14 in relation to the OSI layered model. The SS7 structure is composed of two major functions: the message transfer part (MTP) and a user part. The MTP provides reliable but connectionless transfer of signalling messages between the communicating user functions across the SS7 network. A user part represents a set of functions of a particular type of a user that makes use of the transport capability provided by the MTP, such as the telephone user part (TUP), the data user part (DUP), and the ISDN user part (ISUP).

To provide the functionality equivalent to the OSI network layer, the MTP is enhanced by the signalling connection control part (SCCP). Together this is called the network service part (NSP) that provides the reliable transfer of messages between application parts of SS7. There are no components of the SS7 protocol that correspond to the OSI presentation, session, and transport layers.

The transaction capability application part (TCAP) provides the transaction capabilities in an SS7 network, which can be used as the basis of distributed databases. The TCAP allows an application at one node to invoke execution of a procedure at another node and subsequently obtain the results of such invocation. Built on top of the TCAP, the operation, maintenance, and administration part (OMAP) provides the application protocols and procedures to monitor, coordinate, and control all the network resources. The specific information that an application needs is available in



ASE = Application Service Element
BISUP = B-ISDN User Part
DUP = Data User Part
ISUP = ISDN User Part
MTP = Message Transfer Part
NSP = Network Service Part
OMAP = Operations Maintenance and Administration Part
SCCP = Signalling Connection Part
TCAP = Transaction Capabilities Application Part
TUP = Telephone User Part

Figure 14 : SS7 Protocol Architecture

the application service elements (ASEs). The TCAP also provides all the tools needed by the ASEs for distributed operation between application layers.

The telephone user part (TUP) provides signalling functions to support control of telephone calls on national or international connections. The data user part (DUP) specifies the protocol for inter-exchange signalling for digital data networks. In an ISDN environment, the ISDN user part (ISUP) provides the signalling functions needed to support the basic bearer service, as well as supplementary services, for switched voice and data applications. The ISUP can invoke the services of either the MTP or the SCCP, depending on the function being performed.

The B-ISDN user part (B-ISUP) encompasses signalling functions required to provide switched services and user facilities for voice, video, and data applications in a B-ISDN environment. In the emerging SS7 architecture for B-ISDN, MTP layers 1 and 2 are being replaced by ATM physical, ATM, and SAAL layers. Therefore, two signalling protocol architectures are possible for B-ISDN: the use of B-ISUP over the original MTP; or the use of B-ISUP over MTP3, SAAL, ATM, and a physical facility. Based on the ATM technology, B-ISDN provides a virtual connection transport mechanism for carrying signalling messages as well as all other types of user traffic and management data. Virtual connections can be used as logical signalling links to direct signalling

messages in the network. Consequently, a separate physical signalling network is not needed.

9.2 B-ISDN User Part (B-ISUP)

The B-ISUP is an SS7 protocol which has been specified as the node to node signalling for public ATM networks in support of basic bearer services and supplementary services for B-ISDN applications. The B-ISUP is developed particularly for international applications as a network node interface. The B-ISUP makes use of the services provided by MTP layer 3 which defines a set of transport functions common to all individual signalling links. MTP layer 3 is basically a connectionless protocol that embraces functions summarised in two categories: signalling message handling and signalling network management. Functions related to message handling direct signalling messages to the proper signalling link (i.e., MTP layer 2) or user part. On the basis of predetermined data and information about the status of the signalling network, the signalling network management functions control the current message routing and configuration of the signalling network facilities. In case of a change in the status of the network, these functions are able to control the reconfigurations and other actions to preserve or restore the normal message transfer capability.

The B-ISUP architecture can be viewed as a set of functional blocks, each representing a particular function, as illustrated in Figure 15. ITU-T Recommendations Q.2761 to Q.2764 describe the B-ISUP supporting the Capability Set 1[18] ATM bearer services. The signalling capabilities supported by the B-ISUP are categorised into two classes: internationally applicable class for carriers providing the signalling capabilities over an international boundary, and national use class for national network operators.

9.2.1 B-ISUP Messages

Generally, the B-ISUP messages are the same as those used in ISUP. The B-ISUP performs functions similar to UNI signalling layer 3 in terms of setting up and releasing calls, though it uses somewhat different set of messages. For example, it uses an Initial Address Message (IAM) in place of a SETUP message, but the services are essentially the same. Below is a list of the B-ISUP messages and their description.

- Address Complete (ACM) - sent in the backward direction indicating that all the address signals required for routing the call to the called party have been received.
- Answer (ANM) - sent in the backward direction indicating that the call has been answered.

[18] The Capability Set 1 provides call control for point-to-point single connection calls, but do not efficiently and effectively support multiconnection, multimedia, or multipoint services. In support of point-to-point and point-to-multipoint connections, the Capability Set 2 adds further call control capabilities, additional bearer types, additional connection topologies, and some dynamic rearrangement of these.

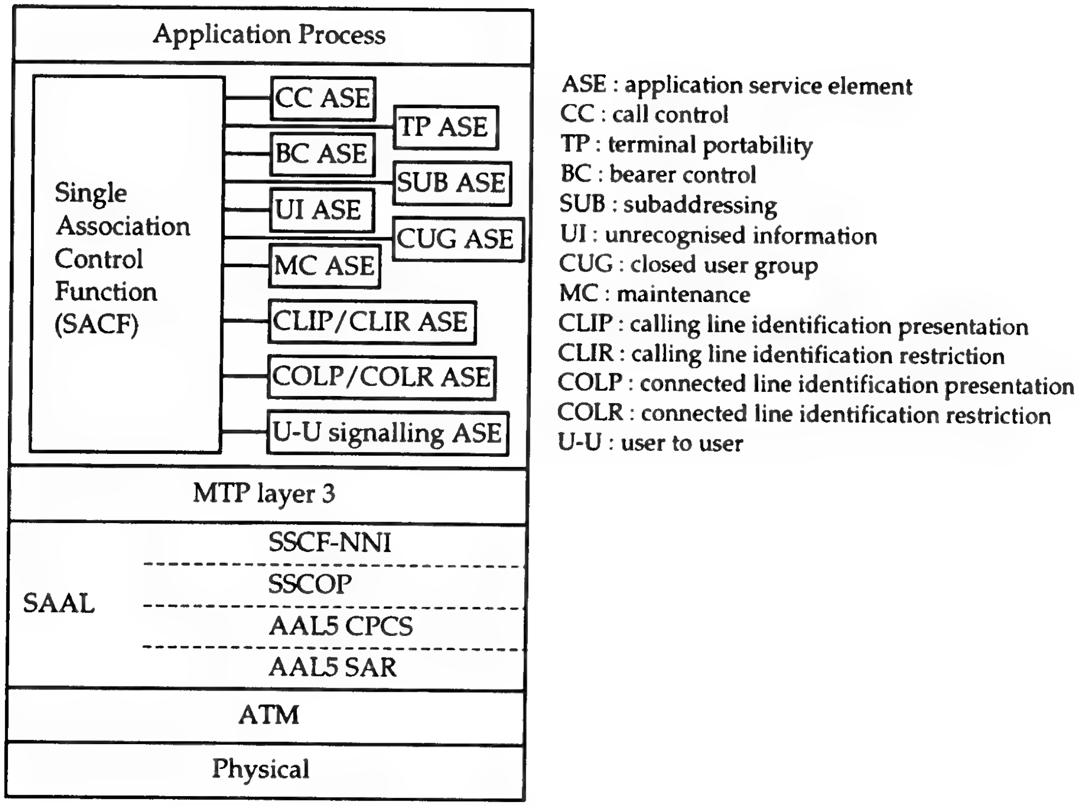


Figure 15 : Functional View of B-ISUP Specification Model

- Blocking Acknowledgment (BLA) - sent in response to a BLO message indicating that the VP is blocked.
- Blocking (BLO) - sent only for maintenance purposes to the exchange at the other end of a VP to cause an engaged condition of that VP for subsequent calls outgoing from that exchange.
- Consistency Check End (CCE) - sent to the exchange at the other end of a VP connection indicating the end of the consistency check sequence and to deactivate the consistency check ATM cell monitoring devices.
- Consistency Check End Acknowledgment (CCEA) - sent in response to a CCE message indicating the result of the consistency check and that the consistency check monitoring device has been activated.
- Consistency Check Request (CCR) - sent for maintenance purposes to the exchange at the other end of a VP connection to verify the consistent and correct allocation of a VPCI to a VP. The test will cause the remote (receiving) exchange to activate an ATM cell monitoring device for the indicated resource.
- Consistency Check Request Acknowledgment (CCRA) - sent in response to a CCR message indicating that the ATM cell monitoring device has been activated for the indicated resource.

- Confusion (CFN) - sent in response to any message if the exchange does not recognise the message or parts of it.
- Call Progress (CPG) - sent in the backward direction indicating that an event has occurred during call setup which should be relayed to the calling party.
- Forward Transfer (FOT) - sent when the outgoing international exchange operator wants the help of an operator at the incoming international exchange.
- IAM Acknowledge (IAA) - sent in response to an IAM message indicating that IAM has been accepted and the requested bandwidth is available in both directions.
- Initial Address (IAM) - sent to initiate seizure of an outgoing VC and to transmit number and other information relating to the routing and handling of the call.
- IAM Reject (IAR) - sent in response to an IAM message indicating call refusal due to resource unavailability.
- Network Resource Management (NRM) - sent to modify network resources associated with a call.
- Reset Acknowledge (RAM) - sent in response to an RSM message indicating that the resources have been released.
- Release (REL) - sent to indicate that the call/connection is being released due to the reason supplied and that the resources are available for new traffic upon receipt of the RLC message.
- Resume (RES) - sent to indicate that the calling or called party, after being suspended, is reconnected.
- Release Complete (RLC) - sent in response to a REL message when the resources of the call/connection concerned have been made available for new traffic.
- Reset (RSM) - sent to release a virtual connection when a REL or RLC message is inappropriate.
- Subsequent Address (SAM) - sent following an IAM message to convey additional called party information.
- Segmentation (SGM) - sent to convey an additional segment of an overlong message.
- Suspend (SUS) - sent to indicate that the calling or called party has been temporarily disconnected.
- Unblocking Acknowledgment (UBA) - sent in response to an UBL message indicating that the resource has been unblocked.
- Unblocking (UBL) - sent to the exchange at the other end of a VPC to cancel the engaged condition of the resource caused by a previously sent BLO message.
- User Part Available (UPA) - sent in response to a UPT message indicating that the user part is available.

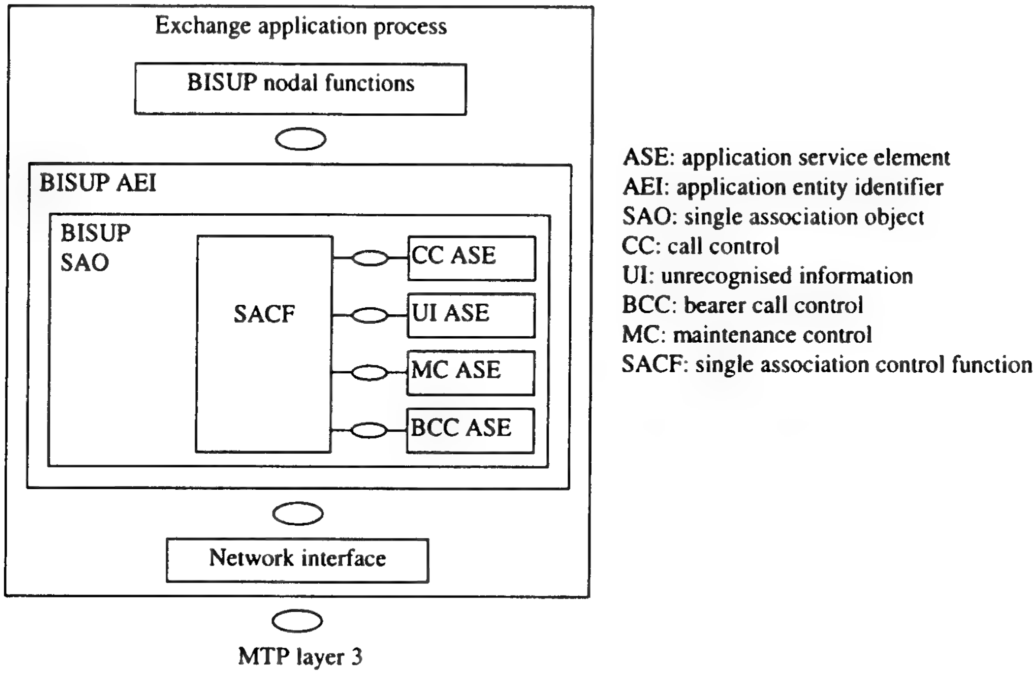


Figure 16 : Generalised Model for B-ISUP Basic Call Application Process

- User Part Test (UPT) - sent to test the status of a user part marked as unavailable for a signalling point.
- User-to-User Information (USR) - used for the transport of user-to-user signalling independent of call control messages.

9.2.2 B-ISUP Signalling Procedures

Recommendation Q.2764 describes the basic B-ISUP signalling procedures for the setup and cleardown of B-ISDN Release 1 network connections in the context of six types of exchanges:

- originating exchange;
- intermediate national exchange;
- outgoing international exchange;
- intermediate international exchange;
- destination exchange;
- incoming international exchange.

The B-ISUP basic call application process contains call control, maintenance, and compatibility management functions. The model is illustrated in Figure 16 where the B-ISUP application entity (B-ISUP AE) provides all the communication capabilities required by the B-ISUP nodal functions. An instance of the B-ISUP AE is created for each signalling association required. A unique signalling identifier (SID) value is used

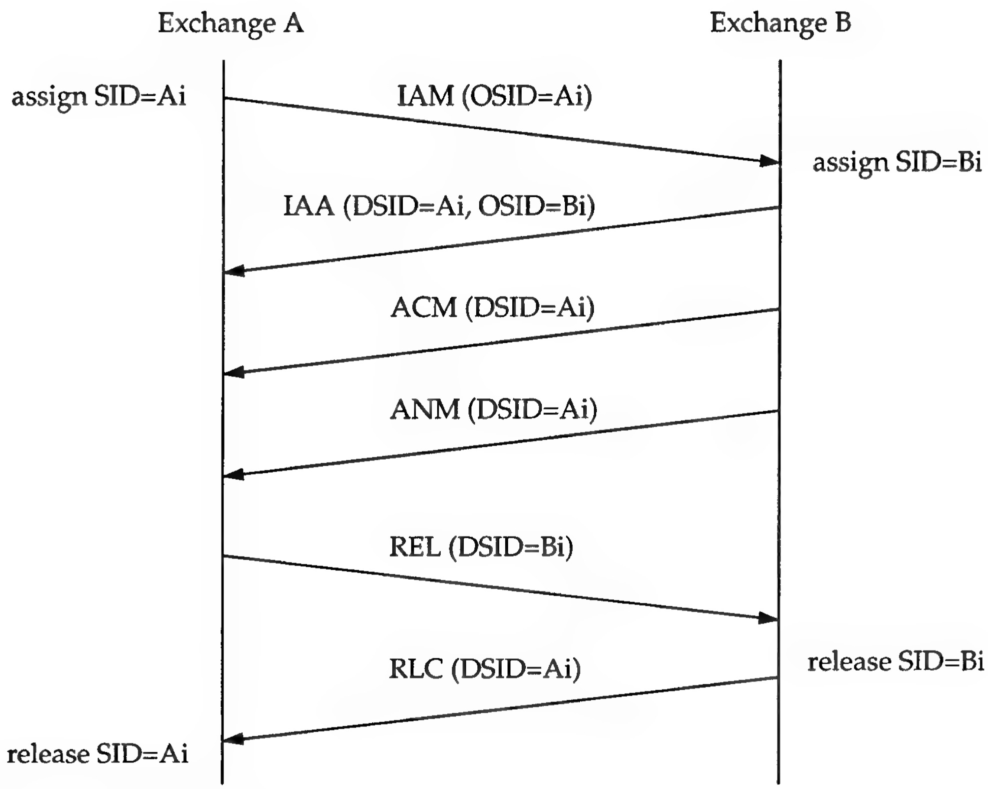


Figure 17 : B-ISUP Signalling Message Flows for a Successful Connection Setup and Release

to identify a particular instance of the B-ISUP AE and all the related signalling messages. The coordination among B-ISUP AEs is provided by the B-ISUP nodal functions. Three types of signalling association are defined: incoming call and connection control; outgoing call and connection control; and maintenance. Upon request, the B-ISUP nodal functions create an instance of the B-ISUP AE which contains a single association object (SAO) of the appropriate type.

The B-ISUP protocol functions are subdivided into five parts: bearer call control (BCC), maintenance control (MC), call control (CC), unrecognised information (UI) application service elements (ASEs), and single association control function (SACF). To communicate with a peer entity in another exchange, the B-ISUP nodal function uses the services of the application entity identifier (AEI). A set of related primitives are passed to ASEs, and the actions are taken at the corresponding ASE. The outcome is contained in a packet and passed to MTP layer 3. At the destination exchange, the message is received by MTP layer 3 and delivered to the respective B-ISUP SAO according to the SID value. Figure 17 depicts an example of a typical signalling session for assigning and releasing the signalling associations. Only the SID parameters of the messages are shown in which OSID and DSID denote the origination signalling identifier and the destination signalling identifier, respectively.

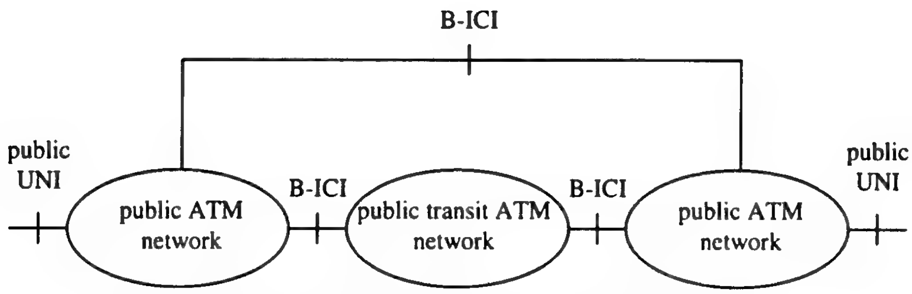


Figure 18 : The B-ICIs Connecting Public ATM Networks

B-ISDN inter-carrier service specific functions
ATM
DS3/E3, SONET/SDH

B-ISDN Inter-Carrier Interface (B-ICI)

ATM
PHY (SONET/SDH)

Network Node Interface (NNI)

Figure 19 : Relation of a B-ICI to the NNI

9.3 B-ISDN Inter-Carrier Interface (B-ICI)

End-to-end national and international ATM services require the interconnection of public carrier networks. The interoperability of multiple ATM based public networks is accomplished by the ATM Forum B-ISDN Inter-Carrier Interface (B-ICI). This specification defines an interface between two different ATM public networks as the demarcation point, to support user services across multiple public carriers. Figure 18 illustrates the B-ICIs connecting two public ATM networks belonging to two different carriers, directly and also via a public transit ATM network, carrying traffic between the generic user-to-network interfaces across the public carriers networks.

The B-ICI specification is based on the ITU-T defined Network Node Interface (NNI) with the addition of DS3 and E3 physical layers. To support Circuit Emulation Service (CES), Frame Relay Service (FRS) and Switched Multimegabit Data Service (SMDS), the B-ICI specification also includes layers above the ATM layer (e.g., AAL, other inter-carrier service specific layers). The relation between the ATM Forum B-ICI and the ITU-T NNI is illustrated in Figure 19.

Similar to the PNNI, the B-ICI is a carrier-to-carrier interface excluding some of the detailed information offered by the PNNI. Public carriers are less likely to share routing information or detailed network maps with their competitors. The initial version of the B-ICI (version 1.1) is designed for the PVC-based inter-carrier services support only. The B-ICI version 2 covers both PVC and SVC capabilities between carrier networks.

The interconnectivity across a B-ICI is obtained by establishing a VCC or VPC at the B-ICI with a QoS suitable to meet the service requirements for end-to-end connections. The ATM cells of a particular service are multiplexed together and passed across the B-ICI to another carrier over DS3/E3, or SONET/SDH facilities. The B-ICI supports multiple services by the ATM network establishing one or more VPCs or VCCs via bilateral agreement from a set of available VPI/VCI values supported on the interface. There is at least one connection for each service supported, and cells belonging to different services are not multiplexed onto the same connection.

10. ATM Signalling in Brief

To establish a switched virtual circuit (SVC) from one point to another across an ATM network, the originating end station sends a SETUP message to the network across its UNI, carrying a number of essential information elements (IEs). These IEs are used by the network and the destination ATM end station to establish a virtual channel connection (VCC). The most important IEs among others are: called party number, ATM traffic descriptor, broadband bearer capability, and QoS parameters [5]. The called party number IE is used by the network to locate the destination node, whereas the latter three are used to define the application requirements for a VCC.

Upon receiving the SETUP message from the originating end station, the switch at the other side of its UNI locates the destination end station, for instance, using some kind of directory service. The IEs in the SETUP message reveal the source traffic characteristics and application, and hence the amount of resources required to support the application. Consequently, a path is selected from the origin node to the destination node. In the PNNI, an end-to-end path is determined by the origin ATM switch based on source routing.

Once a suitable path is found, the origin switch proceeds with the end-to-end connection establishment using some type of internal signalling (PNNI, IISP, B-ISUP or B-ICI). Each node along the path determines whether the required service can be supported, and responds to the origin ATM switch accordingly. If the connection establishment can proceed along the selected path, a SETUP message is constructed and delivered to the destination end station across the UNI.

The destination end station accepts the call by responding with a CONNECT message to the destination ATM switch, which then forwards this information back to the origin ATM switch. Meanwhile, the destination ATM switch sends a CONNECT ACKNOWLEDGE message to the destination end user to indicate the CONNECT message was received and processed. If every switch along the path can support the connection, the origin ATM switch sends a CONNECT message back to the source user, which responds with a CONNECT ACKNOWLEDGE message. The connection is now fully established and ATM cells can start flowing in both directions.

To ensure that the end stations do not send more cells than what they agreed upon, the network monitors the traffic generated by each source at the edge of the network. Each source is expected to stay within the parameters negotiated at the call establishment phase. Nonconforming cells are allowed to enter the network to achieve higher utilisation of network resources, but at a lower priority by marking CLP=1 in the header.

Coupling an enhanced Q.2931 for call control between switches and a dynamic source routing protocol, the PNNI is particularly suitable to the enterprise network environment. The use of the PNNI routing protocol in the public network environment generates some concerns about its pitfalls. These potential problems are being investigated by worldwide research laboratories such as Bellcore and IBM T. J. Watson Research Centre [7].

11. Future Development

The B-ISDN concept has been developed as an approach to integrating a wide range of data, voice, video, and multimedia applications in one integrated network. ATM technology has been accepted as the most flexible, efficient and robust method of supporting such a wide variety of network needs in an integrated manner. Due to the immense size of the standardisation task involved, ATM is being developed by international bodies in different stages (releases or phases), mainly in relation to the development of signalling protocols.

A series of Releases (or Capability Sets) has been defined by the ITU-T and Committee T1, to facilitate graceful evolution to B-ISDN. The B-ISUP Capability Set 1 (CS1) is the SS7 protocol that provides the signalling functions required to support basic bearer services and, where possible, supplementary services for Release 1 B-ISDN applications. B-ISUP CS1 interfaces with the existing SS7 MTP level 3 and provides for ATM connections and call control signalling between public switched network elements for initial broadband services. Increasingly, the signalling protocol is pushed to support additional capabilities as ATM technology becomes more mature. Enhancements have to be made to satisfy the signalling requirements of newly developed network services such as intelligent network applications, mobility management and services, broadband and multimedia applications. While basic signalling provides a component part for the control of network connections, service specific control protocols are able to support more complex multimedia services [7].

The approach recommended for Capability Set 2 (CS2) and beyond to support sophisticated multiconnection, multimedia, and multipoint services is based on call and connection separation. It allows for more efficient call setup in a multipoint and mobile customer environment, and for more flexibility in directing information to destinations at all stages of a call. Recent development of signalling protocols is aimed at supporting enhancements to the basic call control. These functionalities are covered

in B-ISUP CS2, including point-to-multipoint single connection calls, point-to-point multiconnection calls, and point-to-point calls with renegotiation of traffic, edge-to-edge signalling functions (e.g., look ahead), along with several additional capabilities that further define connections. In the distant future, CS3 will add multimedia and broadcast connections.

References

[1] Hiroshi Saito, "Teletraffic Technologies in ATM Networks," Artech House, 1994.

[2] William Stallings, "Data and Computer Communications," Maxwell Macmillan, Third Edition, 1991.

[8] ISO/IEC DIS 13818-6 Information technology -- Generic coding of moving pictures and associated audio information -- Part 6: Extensions for DSM-CC.

[9] ITU-T Recommendation H.245 (03/96) - Control Protocol for Multimedia Communication.

[3] Daniel Minoli, and George Dobrowski, "Principles of Signaling for Cell Relay and Frame Relay," Artech House, 1995.

[4] T. A. Au, and W. D. Blair, "Issues in Asynchronous Transfer Mode Network Addressing for the Australian Department of Defence," DSTO Technical Report DSTO-TR-0595, April 1998.

[5] Raif O. Onvural, "Asynchronous Transfer Mode Networks: Performance Issues," Second Edition, Artech House, 1995.

[6] Anthony Alles, "ATM Internetworking," Cisco Systems Inc. White Paper, May 1995.

[7] Lyndon Y. Ong, "Broadband Network Signaling Directions," Proc. SPIE - Int. Soc. Opt. Eng. (USA), vol. 2917, p. 262-71, 1996.

Appendix : UNI Information Elements

This Appendix contains a full list of information elements defined in UNI Signalling 4.0 Specification. To support signalling, some of these information elements are mandatory for each message type and some are optional. For a complete description of the format, the reader is referred to the original specification.

1. available bit rate (ABR) additional parameters (max 14 octets) - this element is used to specify the set of additional ABR parameters during the call/connection establishment.
2. ABR setup parameters (max 36 octets) - this element specifies the set of ABR parameters during the call/connection establishment.
3. alternative ATM traffic descriptor[19] (max 30 octets) - this element specifies an alternative ATM traffic descriptor for the negotiation of traffic parameters during constant bit rate (CBR) and variable bit rate (VBR) call/connection setup.
4. ATM adaptation layer parameters (max 21 octets) - this information element indicates the requested ATM adaptation layer parameter values for the ATM connection, such as the forward and backward maximum CPCS service data unit (CPCS-SDU) size, value of the MID[20], and user-defined AAL information. Its content is transparent for the network, except for the case of interworking.
5. ATM traffic descriptor (max 30 octets) - a traffic control capability is specified in terms of a set of traffic parameters within this information element such as peak cell rate, sustainable cell rate and maximum burst size in both forward and backward directions.
6. broadband bearer capability (max 7 octets) - it indicates the requested characteristics of a broadband connection oriented bearer service, including traffic type (i.e., constant or variable rate), and timing requirements.
7. broadband high layer information (max 13 octets) - this information element is transferred transparently across an ATM network and is used for compatibility checking of high layer information by an addressed entity.
8. broadband locking shift (max 5 octets) - this information element is employed by the broadband locking shift procedure to activate a new codeset for interpreting following information elements within that message, or until another broadband locking shift information element is encountered.
9. broadband low layer information (max 17 octets) - this is used for compatibility checking of low layer information by an addressed entity including default packet size and window size of a user specified network protocol (e.g., X.25).

---

[19] Not supported in the ATM Forum UNI 3.1 Specification.

[20] The multiplexing identification (MID) field in the AAL 3/4 SAR-protocol data unit (SAR-PDU) structure is 10 bits long and is used to allow multiplexing AAL connections into a single ATM connection.

10. broadband non-locking shift (max 5 octets) - this element is used by the broadband non-locking shift procedure to provide a temporary shift to a specified codeset for the interpretation of the next single information element only.
11. broadband repeat indicator (max 5 octets) - this specifies how repeated information elements shall be interpreted, when included in a message.
12. broadband sending complete (max 5 octets) - this element is to optionally indicate completion of the called party number for compatibility with certain public networks.
13. call state (max 5 octets) - this information element describes the current status of a call or a global interface state.
14. called party number (max 25 octets) - this is to identify the called party of a call using either the E.164 numbering plan or the ATM endsystem address (AESA) format.
15. called party subaddress (max 25 octets) - this is used to identify the subaddress of the called party of a call in the AESA format across a public network which supports only E.164 address.
16. calling party number (max 26 octets) - the origin of a call is identified using either the E.164 numbering plan or the AESA format.
17. calling party subaddress (max 25 octets) - this is to identify a subaddress associated with the origin of a call in the AESA format across a public network which supports only E.164 address.
18. cause (max 34 octets) - this information element identifies the cause and provides diagnostic information in the event of procedural errors.
19. connected number[21] (max 25 octets) - this element is used to identify a number which may be different from the called party number because of changes (e.g., redirection, transfer) during the lifetime of a call.
20. connected subaddress[21] (max 25 octets) - this element is used to identify a subaddress which may be different from the called party subaddress because of changes (e.g., redirection, transfer) during the lifetime of a call.
21. connection identifier (max 9 octets) - this information element describes the ATM connection associated with this call in terms of the virtual path connection identifier (VPCI). The VPI and the VPCI are identical in ATM Release 1[22].
22. connection scope selection[21] (max 6 octets) - this element contains a routing range selected by the calling user within which the call/connection shall be processed and progressed.
23. end-to-end transit delay[21] (max 11 octets) - this element specifies the maximum cumulative end-to-end transit delay acceptable on a per call basis, including the

[21] Not supported in the ATM Forum UNI 3.1 Specification.

[22] Release 1 is primarily aimed at supporting existing telecommunications services, such as telephony and constant bit rate services through ATM.

end user delay (e.g., AAL handling delay) and the forward maximum cell transfer delay.

24. endpoint reference (max 7 octets) - this is to identify the individual endpoints of a point-to-multipoint connection. The first party of a point-to-multipoint call is identified by zero, and the subsequent parties by a non-zero value.

25. endpoint state (max 5 octets) - this element contains the state of an endpoint of a point-to-multipoint connection, such as add party initiated, add party received, drop party initiated, and drop party received.

26. extended quality of service (QoS) parameters (max 25 octets) - this element indicates the individual QoS parameter values acceptable on a per call basis and the cumulative QoS parameter values, including acceptable forward/backward peak-to-peak cell delay variation (CDV), cumulative forward/backward peak-to-peak CDV, and acceptable forward /backward cell loss ratio.

27. generic identifier transport[23] (max 33 octets) - this element carries an identifier for the related standard/application between two users such as digital storage media command & control (DSM-CC) [8], and Recommendation H.245 [9].

28. leaf initiated join call identifier[23] (max 9 octets) - the leaf initiated join (LIJ) call identifier information element is used to uniquely identify a point-to-multipoint call at a root's interface.

29. leaf initiated join parameters[23] (max 5 octets) - this element is used by the root of a point-to-multipoint call to associate options with the call when created such as network join without root notification.

30. leaf sequence number[23] (max 8 octets) - a joining leaf uses this information element to associate a SETUP, ADD PARTY, or LEAF SETUP FAILURE response message with the corresponding LEAF SETUP REQUEST message.

31. minimum acceptable traffic descriptor[23] (max 20 octets) - applicable to CBR, VBR, ABR and unspecified bit rate (UBR) call/connection setup, this information element specifies the minimum acceptable ATM traffic parameters as the lowest values that the user is willing to accept for the call/connection.

32. narrowband bearer capability[23] (max 14 octets) - this element indicates the requested narrowband ISDN bearer capability which is a fallback service of the calling user.

33. narrowband low layer compatibility[23] (max 20 octets) - this element provides a means that could be used for compatibility checking by an addressed entity (e.g., a remote user, an interworking unit, or a high layer function network node addressed by the user).

34. narrowband high layer compatibility[23] (max 7 octets) - this element provides a means that could be used by the remote user for compatibility checking of a fallback service.

---

[23] Not supported in the ATM Forum UNI 3.1 Specification.

35. notification indicator[24] (max length is network dependent) - this element is intended to provide supplementary service[25] notification to a user.
36. progress indicator[24] (max 6 octets) - this element describes an event that has occurred during the lifetime of a call.
37. quality of service parameter (max 6 octets) - this is used to indicate the requested quality of service (QoS) class for each direction of a connection.
38. restart indicator (max 5 octets) - this element is to identify the class of the facility (i.e., virtual channels) to be restarted.
39. transit network selection (max 9 octets) - this is to identify one requested transit network using the U.S. carrier identification codes (CICs) which is only applicable to the U.S. telecommunications environment.

[24] Not supported in the ATM Forum UNI 3.1 Specification.
[25] Examples of the supplementary services are: user-to-user signalling, closed-user group, calling line identification, and call forwarding (i.e., diversion services).

DISTRIBUTION LIST

ATM Signalling in Perspective

T. A. Au

AUSTRALIA

1. DEFENCE ORGANISATION

a. DGC3ID

b. S&T Program

Chief Defence Scientist } shared copy
FAS Science Policy } shared copy
AS Science Corporate Management } shared copy
Director General Science Policy Development
Counsellor Defence Science, London (Doc Data Sheet only)
Counsellor Defence Science, Washington (Doc Data Sheet only)
Scientific Adviser to MRDC Thailand (Doc Data Sheet only)
Director General Scientific Advisers and Trials/Scientific Adviser Policy and Command (shared copy)
Navy Scientific Adviser (Doc Data Sheet and distribution list only)
Scientific Adviser - Army (Doc Data Sheet and distribution list only)
Air Force Scientific Adviser
Director Trials

Aeronautical and Maritime Research Laboratory
Director

Electronics and Surveillance Research Laboratory
Director
Chief of Communications Division
Research Leader Military Information Networks
Head Network Integration Group
Author: T.A. Au

DSTO Library
Library Fishermens Bend
Library Maribyrnong
Library Salisbury (2 copies)
Australian Archives
Library, MOD, Pyrmont (Doc Data sheet only)

c. Capability Development Division
Director General Maritime Development (Doc Data Sheet only)
Director General Land Development (Doc Data Sheet only)
Director General C3I Development (Doc Data Sheet only)

d. Navy

Communications School (Navy), HMAS Cerberus, Hastings, Victoria.

e. **Army**
ABCA Office, G-1-34, Russell Offices, Canberra (4 copies)
School of Signals, Simpson Barracks, Macleod, Victoria.

f. **Air Force**
School of Technical Training, RAAF Base, Wagga, NSW 2651.

g. **Intelligence Program**
DGSTA, Defence Intelligence Organisation

h. **Acquisition and Logistics Program**
PD DEFNET
PD DSDN

i. **Corporate Support Program (libraries)**
OIC TRS, Defence Regional Library, Canberra
Officer in Charge, Document Exchange Centre (DEC), (Doc Data Page only)
DEC requires the following copies of public release reports to meet exchange agreements under their management:
*US Defence Technical Information Centre, 2 copies
*UK Defence Research Information Center, 2 copies
*Canada Defence Scientific Information Service, 1 copy
*NZ Defence Information Centre, 1 copy
National Library of Australia, 1 copy

2. **UNIVERSITIES AND COLLEGES**

Australian Defence Force Academy
Library
Head of Aerospace and Mechanical Engineering
Deakin University, Serials Section (M list), Deakin University Library
Senior Librarian, Hargrave Library, Monash University
Librarian, Flinders University

3. **OTHER ORGANISATIONS**

NASA (Canberra)
AGPS
State Library of South Australia
Parliamentary Library, South Australia

OUTSIDE AUSTRALIA

4. ABSTRACTING AND INFORMATION ORGANISATIONS
INSPEC: Acquisitions Section Institution of Electrical Engineers
Library, Chemical Abstracts Reference Service
Engineering Societies Library, US
Materials Information, Cambridge Scientific Abstracts, US
Documents Librarian, The Center for Research Libraries, US

5. INFORMATION EXCHANGE AGREEMENT PARTNERS
Acquisitions Unit, Science Reference and Information Service, UK
Library - Exchange Desk, National Institute of Standards and Technology, US

SPARES (10 copies)

Total number of copies: 72

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION
DOCUMENT CONTROL DATA

1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)

2. TITLE ATM Signalling in Perspective	3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)
4. AUTHOR(S) T.A. Au	5. CORPORATE AUTHOR Electronics and Surveillance Research Laboratory PO Box 1500 Salisbury SA 5108

6a. DSTO NUMBER	6b. AR NUMBER	6c. TYPE OF REPORT	7. DOCUMENT DATE
DSTO-TR-0671	AR-010-542	Technical Report	May 1998

8. FILE NUMBER	9. TASK NUMBER	10. TASK SPONSOR	11. NO. OF PAGES	12. NO. OF REFERENCES
E8709/4/19 (1)	ADF96/295	DGC3ID	48	7

13. DOWNGRADING/DELIMITING INSTRUCTIONS	14. RELEASE AUTHORITY
	Chief, Communications Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for Public Release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE CENTRE, DIS NETWORK OFFICE, DEPT OF DEFENCE, CAMPBELL PARK OFFICES, CANBERRA ACT 2600

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CASUAL ANNOUNCEMENT Yes

18. DEFTEST DESCRIPTORS

Asynchronous transfer mode, Communications networks, Signal theory (telecommunications)

19. ABSTRACT

Signalling has great impact on the efficient use of network resources, and the services that an Asynchronous Transfer Mode (ATM) network can offer. Importantly, signalling allows the transfer of service-related information in real time between the user and the network, and among network entities to establish, maintain and release end-to-end virtual connections. For each particular signalling function, the corresponding procedures are defined such that the sequence and message format are specific to the network interface across which the exchange of signalling information takes place. This report reviews the signalling architectures for the user-network interface and the network-network interface in both private and public ATM network environments. It is hoped that readers will gain a proper perspective of the signalling protocols involved in ATM networks, and thereby an appreciation of their significant contributions.